

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 09:06:13 ; Search time 7412 seconds  
(without alignments)  
10994.600 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1 gtagataactctcaaaagt.....tttaataataaaaaaaaaa 1992

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapex 1.0

Searched: 288711-seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_ba:  
2: gb\_htg:  
3: gb\_in:  
4: gb\_ov:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pi:  
9: gb\_pi:  
10: gb\_ro:  
11: gb\_ro:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vi:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_mu:  
20: em\_mu:  
21: em\_or:  
22: em\_or:  
23: em\_ov:  
24: em\_ph:  
25: em\_pi:  
26: em\_pi:  
27: em\_pi:  
28: em\_pi:  
29: em\_pi:  
30: em\_pi:  
31: em\_pi:  
32: em\_pi:  
33: em\_pi:  
34: em\_pi:  
35: em\_pi:  
36: em\_pi:  
37: em\_pi:  
38: em\_pi:  
39: em\_pi:  
40: em\_pi:  
41: em\_pi:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1992	100.0	1992	6	AX054819 Sequence
2	1593	80.0	1593	9	AF121863 Homo sapi
3	1540.4	77.3	3038	9	BC005110 Homo sapi
4	1530	76.8	2925	9	BC046520 Homo sapi
5	1525.2	76.6	3616	9	AK095380 Homo sapi
6	1516.6	76.1	3576	9	AK026479 Homo sapi
7	1504	75.5	3145	6	AX512835 Sequence
8	1504	75.5	3145	9	AX000362 Homo sapi
9	1371.4	68.8	2661	9	AY044865 Homo sapi
10	1193.2	59.9	1782	10	BC043328 Mus muscu
11	925.4	46.5	968	9	AF420561 Homo sapi
12	516	25.9	129010	9	AL589666 Human DNA
13	427.2	21.4	451	6	AX331384 Sequence
14	425	21.3	425	6	BD113855 EST and e
15	357	17.9	358	6	BD026500 Sequence
16	172	8.6	278898	2	AC116713 Mus muscu
17	167.2	8.4	224551	2	AC111832 Rattus no
18	163.8	8.2	454	11	G30543 human STS S
19	135.6	6.8	278375	2	AC130093 Rattus no
20	86	4.3	311	6	AX185131 Sequence
21	86	4.3	313	6	AX186466 Sequence
22	84	4.2	299	6	AX188027 Sequence
23	78.6	3.9	580	6	AX387326 Sequence
24	75.4	3.8	167444	9	AP005014 Homo sapi
25	73.4	3.7	183972	2	EX537259 Danio rer
26	72.8	3.7	211118	2	EX284684 Danio rer
27	71.8	3.6	110000	2	PFMAL8P1_12
28	71.2	3.6	125623	3	AC115599 Dictyoste
29	70	3.5	156441	9	AC091869 Homo sapi
30	70	3.5	168885	9	AC091873 Homo sapi
31	69.6	3.5	333321	3	AC116986 Dictyoste
32	69.2	3.5	187310	2	EX005122 Danio rer
33	69	3.5	160411	2	AC135935 Rattus no
34	68.4	3.4	9095	6	AX458545 Sequence
35	68.4	3.4	14422	3	AF466146 Melipona
36	67.4	3.4	14012	3	AE001433 Plasmodiu
37	67.4	3.4	343050	3	PF929353 Plasmodiu
38	67.2	3.4	8056	6	AX599046 Sequence
39	67.2	3.4	195620	2	EX088600 Danio rer
40	67	3.4	141913	9	AL512452 Human DNA
41	66.8	3.4	348600	1	AB063521 Wigglewo
42	66.2	3.3	7461	6	AX346687 Sequence
43	65.8	3.3	161757	2	AC044901 Homo sapi
44	65.6	3.3	6161	6	AX345525 Sequence
45	65.6	3.3	8056	6	AX598900 Sequence

ALIGNMENTS

RESULT 1  
AX054819  
LOCUS  
DEFINITION  
ACCESSION AX054819  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Yue, H., Tang, Y.T. and Azimzai, Y.  
Human sorting nexins  
Patent: WO 0073334-A 3 07-DEC-2000;

AX054819 Sequence 3 from Patent WO0073334.  
AX054819.1 GI:12228268  
1992 bp  
DNA  
linear PAT 13-JAN-2001

Incyte Genomics, Inc. (US)		Location/Qualifiers		Source		
1..1992		/organism="Homo sapiens"		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"				
BASE COUNT		671 a 308 c 372 g 641 t				
ORIGIN						
Query Match		100.0%; Score 1992; DB 6; Length 1992;				
Best Local Similarity		100.0%; Pred. No. 0;				
Matches 1992; Conservative 0;		Mismatches 0; Indels 0; Gaps 0;				
QY	1	GTATGAAACTCCAAAGTATGTAGCTATACCTTTAGTTTATCATTTTCAAAACTGTT 60				
DB	1	GTATGAAACTCCAAAGTATGTAGCTATACCTTTAGTTTATCATTTTCAAAACTGTT 60				
QY	61	TTTCTTTTAAATATGTTTCAATAAACTAGACTGTGTGATATTTGCTATGTA 120				
DB	61	TTTCTTTTAAATATGTTTCAATAAACTAGACTGTGTGATATTTGCTATGTA 120				
QY	121	CTTGATACATTTTGTATTAATTCAGGAACACACAGAAAGGGGAGAAATCATTTGGAAT 180				
DB	121	CTTGATACATTTTGTATTAATTCAGGAACACACAGAAAGGGGAGAAATCATTTGGAAT 180				
QY	181	CAGCAGATAGGTAGCAAAATTAAGGAGTATTCAAAAGTACCAATGGAGGAGCTAT 240				
DB	181	CAGCAGATAGGTAGCAAAATTAAGGAGTATTCAAAAGTACCAATGGAGGAGCTAT 240				
QY	241	GTTGCTTAATATGTTGAGTGAAGTGAAGTGAATTTTATGAAGAGGTATTTGTTGT 300				
DB	241	GTTGCTTAATATGTTGAGTGAAGTGAAGTGAATTTTATGAAGAGGTATTTGTTGT 300				
QY	301	AATGGAAGTATTCCTCAGTGGAGCTGTGAGCACACCTAATCTCCCGAAACCTTGC 360				
DB	301	AATGGAAGTATTCCTCAGTGGAGCTGTGAGCACACCTAATCTCCCGAAACCTTGC 360				
QY	361	TGCATGGAAATATGACATTCATATGACATTTTGTAGGATCCCTCTCGAAAGGAA 420				
DB	361	TGCATGGAAATATGACATTCATATGACATTTTGTAGGATCCCTCTCGAAAGGAA 420				
QY	421	GGAGAAAGAAAGAAATTCCTGTGTTTGTATGATGTTGAAGAAATGATAGAGAGC 480				
DB	421	GGAGAAAGAAAGAAATTCCTGTGTTTGTATGATGTTGAAGAAATGATAGAGAGC 480				
QY	481	AGTTGACACAGCTGAAATTCCTGTGTTTGTATGATGTTGAAGAAATGATAGAGAGC 540				
DB	481	AGTTGACACAGCTGAAATTCCTGTGTTTGTATGATGTTGAAGAAATGATAGAGAGC 540				
QY	541	TGAAATCAAAACTAACAGAAATTCATGCTGCTATAGAGATATCTTGAATTCATGTA 600				
DB	541	TGAAATCAAAACTAACAGAAATTCATGCTGCTATAGAGATATCTTGAATTCATGTA 600				
QY	601	GATCATTTGGCCCAAAATATGAAATTCCTGATGATGATGATGATGATGATGATGAT 660				
DB	601	GATCATTTGGCCCAAAATATGAAATTCCTGATGATGATGATGATGATGATGATGAT 660				
QY	661	TCTACAGAACTCTGACAGATCCAGATGATGATGATGATGATGATGATGATGATGAT 720				
DB	661	TCTACAGAACTCTGACAGATCCAGATGATGATGATGATGATGATGATGATGATGAT 720				
QY	721	TTCCCTTAATGTTGGGAAACACAAATTCCTGATGATGATGATGATGATGATGATGAT 780				
DB	721	TTCCCTTAATGTTGGGAAACACAAATTCCTGATGATGATGATGATGATGATGATGAT 780				
QY	781	GAAATATTAATATCTCTCTGAAACATTAATGAAGAGAAAGGTCTGACATTTGGAAC 840				
DB	781	GAAATATTAATATCTCTCTGAAACATTAATGAAGAGAAAGGTCTGACATTTGGAAC 840				
QY	841	TTTTATCATGATTTTCAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCA 900				
DB	841	TTTTATCATGATTTTCAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCA 900				

QY	901	ACTGACCATTCTCAGCCCTACTCTCAGAAAAACAACAAGACCTTTTCAATCATCTGTTTAA 960
DB	901	ACTGACCATTCTCAGCCCTACTCTCAGAAAAACAACAAGACCTTTTCAATCATCTGTTTAA 960
QY	961	AAATAATCAAAACCGTCTGTAAGAGCTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTCCA 1020
DB	961	AAATAATCAAAACCGTCTGTAAGAGCTCTGTAAGAGCTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTCCA 1020
QY	1021	GGTGATGACTGTAGAGAGCTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTCCA 1080
DB	1021	GGTGATGACTGTAGAGAGCTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTCCA 1080
QY	1081	GTTTCCTCAGCTGGCTTCATCATCTCTTAATGGGAACCTCGAATCCTCTTTAAAAACACCCCT 1140
DB	1081	GTTTCCTCAGCTGGCTTCATCATCTCTTAATGGGAACCTCGAATCCTCTTTAAAAACACCCCT 1140
QY	1141	GGAAATGTATATCTGATTACTATCTTCTAGTGTAAACTAGAACAGCTATTTCCAGGAGCACCG 1200
DB	1141	GGAAATGTATATCTGATTACTATCTTCTAGTGTAAACTAGAACAGCTATTTCCAGGAGCACCG 1200
QY	1201	TTTGTGCTCATCTATAACACTTTCTCAGAGATGCTATATTTCTGTGAAAAAACHACTGAACTCTCG 1260
DB	1201	TTTGTGCTCATCTATAACACTTTCTCAGAGATGCTATATTTCTGTGAAAAAACHACTGAACTCTCG 1260
QY	1261	CTCTCTCCAAGATTAAGCAAAAGGAGCAAAACAGACATTTTGAAGAAATGATGAATTAACAT 1320
DB	1261	CTCTCTCCAAGATTAAGCAAAAGGAGCAAAACAGACATTTTGAAGAAATGATGAATTAACAT 1320
QY	1321	TCCAGATCTGTTAGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCT 1380
DB	1321	TCCAGATCTGTTAGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCT 1380
QY	1381	GTTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTATGTTTATTGGACAT 1440
DB	1381	GTTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTATGTTTATTGGACAT 1440
QY	1441	TCTGATACAGGAACGTGTTTCCAGAGCTCAATAAGGTACAAAGGAAGCTTACCTCTGTGAC 1500
DB	1441	TCTGATACAGGAACGTGTTTCCAGAGCTCAATAAGGTACAAAGGAAGCTTACCTCTGTGAC 1500
QY	1501	ATCTTGGATGTAAACACATTTGGATTGGTATAGATAACCCATTGAAATTTCTGCTGTGCG 1560
DB	1501	ATCTTGGATGTAAACACATTTGGATTGGTATAGATAACCCATTGAAATTTCTGCTGTGCG 1560
QY	1561	AGGGTGGTAGAAATTTACTTTTTGGGTATATCTTAT

```

Db      1981 TAAAAAARAAA 1992
|||||
RESULT 2
AF121863
LOCUS      1593 bp mRNA linear PRI 14-SEP-2001
DEFINITION Homo sapiens sorting nexin 14 (SNX14) mRNA, partial cds.
ACCESSION AF121863
VERSION    AF121863.1 GI:4689265
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1593)
AUTHORS    Teasdale,R.D., Locci,D., Houghton,F., Karlsson,L. and Gleeson,P.A.
TITLE      A large family of endosome-localized proteins related to sorting
            nexin 1
JOURNAL    Biochem. J. 358 (Pt 1), 7-16 (2001)
MEDLINE    21378165
PUBMED     11485546
REFERENCE 2 (bases 1 to 1593)
AUTHORS    Teasdale,R.D., Gleeson,P.A. and Karlsson,L.
TITLE      Identification of eleven novel human sorting nexin molecules. A
            sub-group of the sorting nexin family is associated with the early
            endosomes
JOURNAL    Unpublished
AUTHORS    Teasdale,R.D., Gleeson,P.A. and Karlsson,L.
TITLE      Direct Submission
JOURNAL    Submitted (22-JAN-1999); The R.W. Johnson Pharmaceutical Research
            Institute, 3535 General Atomics Court, San Diego, CA 92121, USA
FEATURES
            source
            1..1593
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            <1..1593
            /gene="SNX14"
            <1..1125
            /gene="SNX14"
            /codon_start=1
            /product="sorting nexin 14"
            /protein_id="AAB27836.1"
            /db_xref="GI:4689265"
            /translation="DFPDPSSERKEKRIIPVFCIDVERNDRAVHPHPSVYVR
            YLEFYVLESKTFBGFAPDAQLPSKRIIGPKNYEFKSKREEFOYLOKLQHPELS
            NSQLADFLSPNGETQFLDKILPDVNLKGIKSVFKLMKEGQHLPEFIMNFINS
            ESPKPSRPBELTILSPSENKKLFNDLPKNNARAEINTERKONQYFWEVMTVEGV
            YDLYMYGVVFPVDPWLHLLMGTRILPKNTLEMTDYVLOCKLQSLFOEHLVSLI
            TLRLDALFCENTSPRLQDKQKAKQTPHEMMNYIPDLVKKLCIGETKYESIRLLFDG
            LQDPVLNKQTYVLLDIVIQELFPFLNKVKQKVTSTSWN"
BASE COUNT 533 a 256 c 294 g 510 t
ORIGIN
Query Match 80.0%; Score 1593; DB 9; Length 1593;
Best Local Similarity 100.0%; Pred. No. 1.6e-271;
Matches 1593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      389 GACATTTTTCAGGATCCCTCTCTGAAGGAAGGAGAGAAAAAGAAAGAAATTCCTGGTTT 448
|||||
Db      1 GACATTTTTCAGGATCCCTCTCTGAAGGAAGGAGAGAAAAAGAAAGAAATTCCTGGTTT 60
|||||
Qy      449 TGATTGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGAACATTGGTCT 508
|||||
Db      61 TGATTGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGAACATTGGTCT 120
|||||
Qy      509 GTCTATAGAGATATCTTGAAATCTATGTAATCAAACTCAACAGAAATTCATGGT 568
|||||
Db      121 GTCTATAGAGATATCTTGAAATCTATGTAATCAAACTCAACAGAAATTCATGGT 180
|||||
Qy      569 GCATTTCTGATGCCAGCTTCCTTCTTAAGAGGATCATTTGGCCCCCAAAATTAATGATTC 628
|||||
Db      181 GCATTTCTGATGCCAGCTTCCTTCTTAAGAGGATCATTTGGCCCCCAAAATTAATGATTC 240
|||||
Qy      629 TTAAGTCAAAGAGGGAAGAGTTCCAAAGAAATATCTACAGAAAATCTTCTGAGCATCCAGAA 688
|||||
Db      241 TTAAGTCAAAGAGGGAAGAGTTCCAAAGAAATATCTACAGAAAATCTTCTGAGCATCCAGAA 300
|||||
Qy      689 CTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGCTGGGGGAAACACAATTT 748
|||||
Db      301 CTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGCTGGGGGAAACACAATTT 360
|||||
Qy      749 CTTGATAAGACTATACACAGATGTAATCTTTGGGAAAATTAATAATCTGTTCTCTGGAAA 808
|||||
Db      361 CTTGATAAGACTATACACAGATGTAATCTTTGGGAAAATTAATAATCTGTTCTCTGGAAA 420
|||||
Qy      809 CTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATTAATCTTGT 868
|||||
Db      421 CTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATTAATCTTGT 480
|||||
Qy      869 GAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTTCAGAA 928
|||||
Db      481 GAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTTCAGAA 540
|||||
Qy      929 AACCAAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGGAATAACA 988
|||||
Db      541 AACCAAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGGAATAACA 600
|||||
Qy      989 CAGAGAAAGCAAAATCAGAAATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATGAT 1048
|||||
Db      601 CAGAGAAAGCAAAATCAGAAATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATGAT 660
|||||
Qy      1049 TACCTGATGATGATGAGACGGGTAGTTTTCAGGTTTCTGAGTGGTTCATCATCTCTTA 1108
|||||
Db      661 TACCTGATGATGATGAGACGGGTAGTTTTCAGGTTTCTGAGTGGTTCATCATCTCTTA 720
|||||
Qy      1109 ATGGGAATCGAATCCTCTTTTAAACACCTGGAATGATATGATGATGATGATGATGAT 1168
|||||
Db      721 ATGGGAATCGAATCCTCTTTTAAACACACCTGGAATGATGATGATGATGATGATGAT 780
|||||
Qy      1169 TGTAACCTAGAACAGCTATTTTACAGGAGCAGCGTTTGGTCTCAGTCATAACACTTCTCAGA 1228
|||||
Db      781 TGTAACCTAGAACAGCTATTTTACAGGAGCAGCGTTTGGTCTCAGTCATAACACTTCTCAGA 840
|||||
Qy      1229 GATGCTATATTCTGTGAAAACACAGCACTCGCTCTCTCCAAAGATAAGCAAAAAGGAGCA 1288
|||||
Db      841 GATGCTATATTCTGTGAAAACACAGCACTCGCTCTCTCCAAAGATAAGCAAAAAGGAGCA 900
|||||
Qy      1289 AAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATTGGT 1348
|||||
Db      901 AAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATTGGT 960
|||||
Qy      1349 GAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGGTGGCTTACAGCAACCAAGTACTC 1408
|||||
Db      961 GAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGGTGGCTTACAGCAACCAAGTACTC 1020
|||||
Qy      1409 AACAGAGCTGACTTATGTTTATTTGGACATTTGATATGATGATGATGATGATGATGAT 1468
|||||
Db      1021 AACAGAGCTGACTTATGTTTATTTGGACATTTGATATGATGATGATGATGATGATGAT 1080
|||||
Qy      1469 AATAAGGTACAAAAGGAGTTACCTCTGTGACATCTTGGATGTAAACACATCTGGATTTGGT 1528
|||||
Db      1081 AATAAGGTACAAAAGGAGTTACCTCTGTGACATCTTGGATGTAAACACATCTGGATTTGGT 1140
|||||
Qy      1529 ATAGAAATACCCATTGAAATTTCTGCTGGGAGGGGTGGTAAATTTACTTTTTTGGGT 1588
|||||
Db      1141 ATAGAAATACCCATTGAAATTTCTGCTGGGAGGGGTGGTAAATTTACTTTTTTGGGT 1200
|||||
Qy      1589 ATATTCTTTATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1648
|||||
Db      1201 ATATTCTTTATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
|||||
Qy      1649 TAAAGACTAAACAAAATTAATGATTAAGAGTATGAGTCTCATAGTCTTCTTCTTCTTCT 1708
|||||

```













AUTHORS Daemen,M.J., Cleutjens,C.B. and Zaman,G.J.  
 TITLE Markers of unstable atherosclerotic plaques  
 JOURNAL Patent: WO 02062839-A 4 15-AUG-2002;  
 Universiteit Maastricht (NL)

FEATURES Location/Qualifiers  
 source 1..3145

BASE COUNT 985 a 570 c 659 g 931 t  
 ORIGIN /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

Query Match 75.5%; Score 1504; DB 6; Length 3145;  
 Best Local Similarity 99.5%; Pred. No. 7e-256;  
 Matches 1529; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY	143	TTTCAAGAACACACAGAAAAGGGAGAAATCATTTTGAATCAGCAGAAATAGGTAGCAAAATT	202
Db	1598	TTTCGAAACACACAGAAAAGGGAGAAATCATTTTGAATCAGCAGAAATAGGTAGCAAAATT	1657
QY	203	AAAGGAGTATTCMAAGTACCACATGGAGGAGCTATGTTGCCTAATATATGGTGTAGCT	262
Db	1658	AAAGGAGTATTCAGAAATACCAATGGAGGAGCTATGTTGCCTAATATATGGTGTAGCT	1717
QY	263	GAAGGTGAAGATGATTTTATTCGAAGAGGTATTTGTTAAATGGAAGATGATTTCTCCAGTG	322
Db	1718	GAAGGTGAAGATGATTTTATTCGAAGAGGTATTTGTTAAATGGAAGATGATTTCTCCAGTG	1777
QY	323	GAGGCTGTGAGCACACCTAATACTCCCGAAACCTTGTCTGCATGGAAATATAGCAATCCA	382
Db	1778	GAGGCTGTGAGCACACCTAATACTCCCGAAACCTTGTCTGCATGGAAATATAGCAATCCA	1837
QY	383	TATGTAGACTTTTTCAGAGATCCCTCTCTGAAAGAGGAGAAAGAAAGAAATTCCT	442
Db	1838	TATGTAGACTTTTTCAGAGATCCCTCTCTGAAAGAGGAGAAAGAAAGAAATTCCT	1897
QY	443	GTGTTTTGTATTGATCTTGAAGAAATGATAGAGCAGAGTGGACAGGCGCTCAACAT	502
Db	1898	GTGTTTTGTATTGATCTTGAAGAAATGATAGAGCAGAGTGGACAGGCGCTCAACAT	1957
QY	503	TGGTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATT	562
Db	1958	TGGTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATT	2017
QY	563	CATGGTGCAATTCCTGATGCCAGCTTCCTTTCTAAGAGATCATTTGGCCCCAAAAATTAT	622
Db	2018	CATGGTGCAATTCCTGATGCCAGCTTCCTTTCTAAGAGATCATTTGGCCCCAAAAATTAT	2077
QY	623	GAATTTCTTAAGTCAAGAGGGAGAGTTCCAGAAATATCTACAGAACTTCTGCAGCAT	682
Db	2078	GAATTTCTTAAGTCAAGAGGGAGAGTTCCAGAAATATCTACAGAACTTCTGCAGCAT	2137
QY	683	CCAGAACTGATTAATAGTCAACTTCTGGCAGACTTTCTTCCCTAATAGTGGGGAACA	742
Db	2138	CCAGAACTGATTAATAGTCAACTTCTGGCAGACTTTCTTCCCTAATAGTGGGGAACA	2197
QY	743	CAATTTCTTGATAGATATCTACAGATGTAATCTTTGGGAAATATATAAAATCTGTCTCT	802
Db	2198	CAATTTCTTGATAGATATCTACAGATGTAATCTTTGGGAAATATATAAAATCTGTCTCT	2257
QY	803	GGAAACTAATGAAGAGAAAGTCAAGATTTGGAACTTTTATCATGAATTCATTAAT	862
Db	2258	GGAAACTAATGAAGAGAAAGTCAAGATTTGGAACTTTTATCATGAATTCATTAAT	2317
QY	863	TCTTGTGAGTCTCAAGAGCTTAACCAAGTAGACAGAACTGACCACTTCTGAGCCCTACT	922
Db	2318	TCTTGTGAGTCTCAAGAGCTTAACCAAGTAGACAGAACTGACCACTTCTGAGCCCTACT	2377
QY	923	TCAGAAAACCAACAGAGCTTTTCAATGATCTCTTTTAAATAATATCAAAACCGTCTGAA	982
Db	2378	TCAGAAAACCAACAGAGCTTTTCAATGATCTCTTTTAAATAATATCAAAACCGTCTGAA	2437
QY	983	AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGGAGGTGATGATCTGTAGAGGAGTC	1042

RESULT 8  
 AK000362  
 LOCUS Homo sapiens cDNA FLJ20355 fis, clone HEP15804, highly similar to  
 DEFINITION AF121863 Homo sapiens sorting nexin 14.  
 ACCESSION AK000362  
 VERSION AK000362.1 GI:7020397  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,  
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,  
 Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,  
 Nakamura,Y., Isogai,T. and Sugano,S.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 3145)  
 Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
 Shibahara,T., Tanaka,T. and Nakamura,Y.  
 Direct Submission  
 Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,

2438 AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGGAGGTGATGATCTGTAGAGGAGTC 2497  
 1043 TATGATTACCTGATGATGTAGGACGGGTAGTTTCCAGGTTCCCTGACCTGGCTTTCATCAT 1102  
 2498 TATGATTACCTGATGATGTAGGACGGGTAGTTTCCAGATTCCTGACCTGGCTTTCATCAT 2557  
 1103 CTCTTAATGGAACTCGAATCTCTTTAAAAACACCCCTCGGAAATGTATCTGATTACTAT 1162  
 2558 CTCTTAATGGAACTCGAATCTCTTTAAAAACACCCCTCGGAAATGTATCTGATTACTAT 2617  
 1163 CTTCAAGTAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCCTCCTACATAACACTT 1222  
 2618 CTTCAAGTAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCCTCCTACATAACACTT 2677  
 1223 CTTCAAGATCTATATCTTGTGAAACACCTGACCTCGCTCTCTCCCAAGATAAGCAAAA 1282  
 2678 CTTCAAGATCTATATCTTGTGAAACACCTGACCTCGCTCTCTCCCAAGATAAGCAAAA 2737  
 1283 GAGCAAAAACAGACTTTTGAAGAAATGATGAATACATTCCAGATCTGTTAGTCAAGTGT 1342  
 2738 GAGCAAAAACAGACTTTTGAAGAAATGATGAATACATTCCAGATCTGTTAGTCAAGTGT 2797  
 1343 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGTATGGCTTACAGCAACCA 1402  
 2798 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGTATGGCTTACAGCAACCA 2857  
 1403 GTACTCAACAGCAGCTGACTATGTTTATTTGGACATTTGTGTATCAGGAACTGTTTCCA 1462  
 2858 GTACTCAACAGCAGCTGACTATGTTTATTTTGGACATTTGTGTATCAGGAACTGTTTCCA 2917  
 1463 GAGCTCAATGAAGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAAACACTTGA 1522  
 2918 GAGCTCAATGAAGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAAACACTTGA 2977  
 1523 TTTGGTATAGATAAACCCATTGAAATTTCTGCTGTGCGAGGGTGTAGAAAATTTACTTTT 1582  
 2978 TTTGGTATAGATAAACCCATTGAAATTTCTGCTGTGCGAGGGTGTAGAAAATTTACTTTT 3037  
 1583 TTGGGTATATCTTATATATATTTATGTATACATCCCTGTCTGAAATTTTATTTTGT 1642  
 3038 TTGGGTATATCTTATATATATTTATGTATACATCCCTGTCTGAAATTTTATTTTGT 3096  
 1643 TTTTAAATAAGACTTAACACAACTTAATGATTAATA 1678  
 3097 TTTTAAATAAGACTTAACACAACTTAATGATTAATA 3131

Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna1@ims.u-tokyo.ac.jp,  
Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)  
NEBO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

## FEATURES

## source

Location/Qualifiers  
1..3145  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP15804"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_lib="HEP"  
/note="cloning vector pME18SPL3"

## misc\_feature

1..3145  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14"

BASE COUNT 985 a 570 c 659 g 931 t

## ORIGIN

Query Match 75.5%; Score 1504; DB 9; Length 3145;  
Best Local Similarity 99.5%; Pred. No. 7e-256;  
Matches 1529; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
143 TTCAGGACACACAGAAAGGGAGATCATTTGGAAATCAGCAGAAATGAGTACCAAAAT 202  
1598 TTTCGGAACACACAGAAAGGGAGATCATTTGGAAATCAGCAGAAATGAGTACCAAAAT 1657  
203 AAAGAGTATTCAAAGTACCACAAATGAGGAGCTATGTTGCCATAATTAGGTAGCT 262  
1658 AAAGAGTATTCAGAGTACCACAAATGAGGAGCTATGTTGCCATAATTAGGTAGCT 1717  
263 GAAGTGAAGATGATTTTATGAAAGAGTATTTGTAATGGAAGATGATTTCCAGTG 322  
1718 GAAGTGAAGATGATTTTATGAAAGAGTATTTGTAATGGAAGATGATTTCCAGTG 1777  
323 GAGGCTGTGACACACCTTAATCTCCCGAAACCTTCTGTCAGTGGAAATAGCATTTCCA 382  
1778 GAGGCTGTGACACACCTTAATCTCCCGAAACCTTCTGTCAGTGGAAATAGCATTTCCA 1837  
383 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGAGAGAAAGAAAGAAATTCCT 442  
1838 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGAGAGAAAGAAAGAAATTCCT 1897  
443 GTGTTTCTGATTGATGTTGAAGAAATGATAGAGAGAGTGGACACGAGCTGGAACAT 502  
1898 GTGTTTCTGATTGATGTTGAAGAAATGATAGAGAGAGTGGACACGAGCTGGAACAT 1957  
503 TGGTCTGCTATAGAGATATCTTGAAATCTATGATCTTGAAATCAAAACTTAACAGATTT 562  
1958 TGGTCTGCTATAGAGATATCTTGAAATCTATGATCTTGAAATCAAAACTTAACAGATTT 2017  
563 CATGGTGATTTCTGTGATGCCAGCTTCTCTTAAGAGATCATTTGCCCCCAAAATTTAT 622  
2018 CATGGTGATTTCTGTGATGCCAGCTTCTCTTAAGAGATCATTTGCCCCCAAAATTTAT 2077  
623 GAATTTCTTAAGTCAAAGAGGAGAGTTCGAAGATATCTACAGAAATTTCTGCAGCAT 682  
2078 GAATTTCTTAAGTCAAAGAGGAGAGTTCGAAGATATCTACAGAAATTTCTGCAGCAT 2137  
683 CCAGAACTGAGTAATAGTCAATCTTGGCAGACTTTTCTCCCTAATGTTGGGGAACA 742  
2138 CCAGAACTGAGTAATAGTCAATCTTGGCAGACTTTTCTCCCTAATGTTGGGGAACA 2197  
743 CAATTTCTGTATAGATCTACAGATCTAATCTTGGGAAATATATAAATCTGTTCCT 802  
2198 CAATTTCTGTATAGATCTACAGATCTAATCTTGGGAAATATATAAATCTGTTCCT 2257

803 GGAAACTTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAAT 862  
2258 GGAAACTTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAAT 2317  
863 TCTTGTAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGAGCAATTTCTCAGCCCTACT 922  
2318 TCTTGTAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGAGCAATTTCTCAGCCCTACT 2377  
923 TCAGAAACACACAGAGCTTTTCAATCATCTGTTTAAATAATGCAAAACCGTCTGAA 982  
2378 TCAGAAACACACAGAGCTTTTCAATCATCTGTTTAAATAATGCAAAACCGTCTGAA 2437  
983 AATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGTGATGATCTAGTAAAGGAGTC 1042  
2438 AATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGTGATGATCTAGTAAAGGAGTC 2497  
1043 TATGATTAACCTGATGATGATGAGGAGCTTTTCCAGGTTCTGACTGCTTCATCAT 1102  
2498 TATGATTAACCTGATGATGATGAGGAGCTTTTCCAGGTTCTGACTGCTTCATCAT 2557  
1103 CTCCTTAATGGGAACCTGGAATCTCTTTTAAACACCCCTGGAATGATATCTACTGATCTAT 1162  
2558 CTCCTTAATGGGAACCTGGAATCTCTTTTAAACACCCCTGGAATGATATCTACTGATCTAT 2617  
1163 CTTCACTGTAACTAGAACAGCTATTTCCAGGACACCGTTTGGTCTCTCATCTATAACACTT 1222  
2618 CTTCACTGTAACTAGAACAGCTATTTCCAGGACACCGTTTGGTCTCTCATCTATAACACTT 2677  
1223 CTCAGAGATGCTATATTTCTGTGAAACACCTGAACTCTCTCTCCAGAGTAAGCAAAA 1282  
2678 CTCAGAGATGCTATATTTCTGTGAAACACCTGAACTCTCTCTCCAGAGTAAGCAAAA 2737  
1283 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTTACATCCAGATCTGTTAGTCAAGTGT 1342  
2738 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTTACATCCAGATCTGTTAGTCAAGTGT 2797  
1343 ATTGTGTAGAAACCAAGTATGAAGCATCAGCTCTCTGTTGGTGGCTTACAGCAACA 1402  
2798 ATTGTGTAGAAACCAAGTATGAAGCATCAGCTCTCTGTTGGTGGCTTACAGCAACA 2857  
1403 GTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATACAGGAACTGTTTCCA 1462  
2858 GTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATACAGGAACTGTTTCCA 2917  
1463 GAGCTCAATAAGGTACAAAGAGAGTTACCTCTGTGACATCTTGGATGTAAACACTTGG 1522  
2918 GAGCTCAATAAGGTACAAAGAGAGTTACCTCTGTGACATCTTGGATGTAAACACTTGG 2977  
1523 TTTGGTATAGATAACCCATTTGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 1582  
2978 TTTGGTATAGATAACCCATTTGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 3037  
1583 TTTGGTATATTTCTTATATATATATATGATCATCGCTGCTGAAATTTTATGTTATTTTGT 1642  
3038 TTTGGTATATTTCTTATATATATATATGATCATCGCTGCTGAAATTTTATGTTATTTTGT 3096  
1643 TTTTAAATAAGACTAACCAAACTTTAATGATTAA 1678  
3097 TTTTAAATAAGACTAACCAAACTTTAATGATTAA 3131

## RESULT 9

AY044865

## LOCUS

AY044865 Homo sapiens sorting nexin 14 (SNX14) mRNA, complete cds.

## DEFINITION

AY044865

## ACCESSION

AY044865.1 GI:15529063

## VERSION

KEYWORDS

## SOURCE

ORGANISM

REFERENCE

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 2661)



**AUTHORS** Hong, W.  
**TITLE** The complete coding region of SNX14  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 2661)  
**AUTHORS** Hong, W.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (11-JUL-2001) Membrane Biology Laboratory, Institute of Mol & Cell Biol, 30 Medical Drive, Singapore 117609  
**FEATURES** Location/Qualifiers  
**source** 1. .2661  
`/organism="Homo sapiens"`  
`/mol_type="mRNA"`  
`/db_xref="taxon:9606"`  
1. .2661  
`/gene="SNX14"`  
1. .2661  
`/gene="SNX14"`  
`/note="contains PX domain"`  
`/codon_start=1`  
`/product="sorting nexin 14"`  
`/protein_id="AAK97796.1"`  
`/db_xref="GI:15529064"`  
`/translation="MQOKLQRLRLDVGREICROYPLFCILLCLISASLLLNRYIHI  
LMIFWFGAVGVTFYCSLGDLSLLPNFTIKTKPKQLGLQELFQGHSCAVCGKVK  
KRRPSSLIDENYQWMDLQISKVDASLSEVDIPSIITKGLAKAAKHIEVIVKARQ  
VKNTEFLQAALEEEYFELHVARSRDELHYRLKTELLFYVLPKATDCRSITLL  
IRILSSVFLPSLDLADPTVNHILLIPIDDSPPEKATEPASPLVPLOKPAEPRN  
KPSVUKLEHQRIREQDILLFRPMFLKQEGAVHVLQCLTWEEFNDRILRLPELSENDE  
MUSLHELOKIYQYICLDSIDIKRPDPFIVEIQRIAGPYIDVVKLGQTMKCLFEAY  
EHLVSLLENVTFPMFCHSDSEYFQLLRGAASPTRNSKLANRTQKSGESGISRIQSKI  
KGVPTKTMGKALPNVGAEGDDTIHEGIVVMEDDSPEZAVSTPNTNLAWVKIS  
IPVDPFEDPSSSRERKERRIPVPCIDIVERNDRRAVGHPEHMSVYRRVLYEVLVESK  
LTFPHGAFPDQALPSKRIIGPKNYEPLKSRBEFOYLOKLLQHPLENSOLLADPLS  
ENGETQFDLLIDLPDNLGKIKSVPEKLMKQKHLEPIFNFINSCSPKPKPSRP  
PILFTLSPSTENKCLFNDIPKNNANAENTERNQNYFMENVTFEGVDYLYMYGVR  
VFQDPWLHMLHGLTRILFLNTLWYTDYLVCKLQELQOEHLVSLITLLRDPAIFCE  
NTPRSLQDKQKAGKOTFSEMMNYIPDLLVKCIGIBETKYESIRLLFDGLQQPVLNKLQ  
TYVLLDIVIQLELPENLKYQKEVTSVTSNM"`  
**BASE COUNT** 855 a 490 c 534 g 782 t  
**ORIGIN**  
Query Match 68.8%; Score 1371.4; DB 9; Length 2661;  
Best Local Similarity 96.28; Pred. No. 1.9e-232;  
Matches 1405; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
  
Qy 53 AAACGTGTTTTCTTTTTTAAATAATGTTTTTCAATAACTAGACTGTTGTGATATTT 112  
Db 1201 AATGTAATTACTCTATGTTCTGCCATGATGATGATTTTCAGACAACTTTTAAGAGGT 1260  
  
Qy 113 GCTATGACTTGATACATTTTGTGTTTAATATTCAGGAACACACAGAAAGGGGAGATCA 172  
Db 1261 GCAGAAATCACCAACACGCAATTCAAAATTGAAACAGGAACACACAGAAAGGGGAGATCA 1320  
  
Qy 173 TTGTGAATCAGCAGAAATAGTAGCAAAATTAAGGAGTATTCAAAAGTACCACCAATGGAG 232  
Db 1321 TTGTGAATCAGCAGAAATAGTAGCAAAATTAAGGAGTATTCAAAAGTACCACCAATGGAG 1380  
  
Qy 233 GGAGCTATGTTGCTTAATATGTTGTTAGCTGAAGGTGAAGATGATTTTATTTGAAGAAGGT 292  
Db 1381 GGAGCTATGTTGCTTAATATGTTGTTAGCTGAAGGTGAAGATGATTTTATTTGAAGAAGGT 1440  
  
Qy 293 ATTGTTGTAATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCCGA 352  
Db 1441 ATTGTTGTAATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCCGA 1500  
  
Qy 353 AACCTTGCTGCATGGAAAAATAGCATTTCCATATGTAGACTTTTTTGAAGGATCCCTCTCT 412  
Db 1501 AACCTTGCTGCATGGAAAAATAGCATTTCCATATGTAGACTTTTTTGAAGGATCCCTCTCT 1560  
  
Qy 413 GAAAGGAAGGAAAAAGAAAGAAATTCCTGCTGTTTTGTATTGATGTTGAAGAAATGAT 472  
Db 1561 GAAAGGAAGGAAAAAGAAAGAAATTCCTGCTGTTTTGTATTGATGTTGAAGAAATGAT 1620



```
RESULT 10
BC043328
LOCUS
DEFINITION
  BC043328 1782 bp mRNA linear ROD 16-APR-2003
  Mus musculus RIKEN cDNA C330035N22 Gene, mRNA (cDNA clone MGC:49424
  IMAGE:5400157), complete cds.
ACCESSION
  BC043328
VERSION
  BC043328.1 GI:27694048
KEYWORDS
  MGC.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Mus musculus
REFERENCE
  1 (bases 1 to 1782)
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 1782)
  Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,
  Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
  Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
  Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
  Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
  Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
  Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S.,
  Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
  Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
  McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
  Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
  Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
  Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
  Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
  Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
  Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
  Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,
  Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
  human and mouse cDNA sequences
  Generation and initial analysis of more than 15,000 full-length
  human and mouse cDNA sequences
  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  22388257
  12477932
  2 (bases 1 to 1782)
  Strausberg, R.
  Direct Submission
  Submitted (09-JAN-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: The Cepko Laboratory
  CDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
  BC Cancer Agency, Vancouver, BC, Canada
  info@bcgsc.bc.ca
  Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
  Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
  Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
  Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
  Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline
  Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott,
  Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,
  George Yang, Scott Zuyderduyn, Marco Marra.
  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: IRAC Plate: 86 Row: P Column: 11
  This clone was selected for full length sequencing because it
  passed the following selection criteria: Hexamer frequency ORF
  analysis.
  Location/Qualifiers
  1..1782
  /organism="Mus musculus"
  /mol_type="mRNA"
  /db_xref="taxon:10090"
  /clone="MGC:49424 IMAGE:5400157"
  FEATURES
  source
    /tissue_type="Eye, retina, mouse strain C57Bl/6"
    /clone_lib="NIH_MGC_94"
    /lab_host="DH10B"
    /note="Vector: pCMV-SPORT6"
  1..1782
    /gene="C330035N22Rik"
    /note="synonym: B830022K16"
    /db_xref="LocusID:244962"
    /db_xref="MGI:2443156"
    /codon_start=1
    /product="C330035N22Rik protein"
    /protein_id="AAH43328.1"
    /db_xref="GI:27694048"
    /translation="MRICLIFRSQVSTKRGESFGISGRISKIKGFKSTMEGAVLPNYG
    VAEGDDTIERGIVMEDDSVEAVSTQSTPRNLAANKISIPYVDFDFDPSSEKKEK
    BRIPYCDVERNDRAVHEPHEPVSRYRLEFVLEFVLESLKLEFEGTQFDALPKRI
    IGPKVETLKRKEEFQVLOKLVHQPESLSOLLADFLSPNGGGTQFLDKLPLDNL
    GKLIKSGKLMKEKQHLPEFIPSPINSCSPKPKSPRLPILTSIPSENNKLPND
    LFNKNAENATERNKQNYFMEVNTDGVYDLMYGVRFVFPDWDHLLHMGTRIL
    FKNTLEMTDYLOKLEQLFQEHRLVSLITLLRDAIFCENTEPESLQDKQKQKQTF
    EBMNYIPDLIVKICIGETKYESIRLLPDLGLQQLPVNLKQLTYVLLDIVIQELFELNK
    VQKEATSMTSMW"
  BASE COUNT 567 a 339 c 380 g 496 t
  ORIGIN
    Query Match 59.9%; Score 1193.2; DB 10; Length 1782;
    Best Local Similarity 86.5%; Pred. No. 6e-201;
    Matches 1374; Conservative 0; Mismatches 183; Indels 31; Gaps 4;
  79 TGTGTTTCAATAAAGTAGACTGTTGTGATAATTTGCTATGACTGATGATACATTTTGT 138
  182 TGGCATTGCTATAGTAGTCGGTGGTAAATTTGCTATGACTGATGATGATTTTGT 241
  139 AATATTCAGGACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGTAGCA 198
  242 AATATTTAGAACACACAAAAAGAGAGAGATCGTTTGGAAATCAGCAGAAATAGTAGCA 301
  199 AATTAAAGAGATATTCAAAAGTACCACAAATGAGGAGGAGCTATGTTGCCTAATATTGGTGT 258
  302 AATTAAAGCGGATTCAGAGATACGCAATGAGGAGGAGCTGCTGCTGCTAATTAACGGGT 361
  259 AGCTGAAGGTGAAGATGATTTTATGAAGAGATTTGTTGTAATGGAAGATGATTTCC 318
  362 GGGTGAAGCGGAAGATGACTTTTATGAAGAGGATTTGGTAAATGAGGATGACTTCC 421
  319 AGTGAAGGTGAGCAGACACCTAATCTCCCGAACCTTGTGCTGATGAGAAATAGCAT 378
  422 AGTAGAAGCTGTGAGCAGACACCTTAACTCTCGAACCTTGTGCTGAGAAATAGCAT 481
  379 TCCATATGTAGACTTTTGTGAGGATCCCTCTCGAACGAGGAGAGAGAGAGAGAGAT 438
  482 TCGTATGATGACTTTTGTGAGATCCCTCTCGAACGAGGAGAGAGAGAGAGAT 541
  439 TCGTGTGTTTGTATGATGTTGAAAGAAATGATAGAGAGAGAGAGAGAGAGAGAT 498
  542 TCGTGTGTTTGTATGATGTTGAAAGAAATGATAGAGAGAGAGAGAGAGAGAGAT 601
  499 ACATTGGTCTGTATAGAGATATCTTGAATTTCTTGAATCTTGAATCTTGAATCTTGA 558
  602 GCATTGGTCTGTATAGAGATATCTTGAATTTCTTGAATCTTGAATCTTGAATCTTGA 661
  559 ATTTCATGCTGATTTCTGATGCTCCAGCTCCCTCTTAAGAGAGATCATTTGGCCCAAAA 618
  662 ATTTCATGCTGATTTCTGATGCTCCAGCTCCCTCTTAAGAGAGATCATTTGGCCCAAAA 721
  619 TTATGAATTTCTTAAGTCAAGAGGAGAGATTTCCAAAGATATCTTACAGAAATCTTCTGCA 678
  722 CTATGATGTTTGAAGTCGAGAGAGAGAGATTTCCAGAGATATCTTCCAGAAATCTTCTGCA 781
  679 GCATCCAGAACTGAGTAATAGTCAACTTCTGCAAGCTTTCTTCCCTTAATGTTGGGGA 738
```

Db 782 GCACCGGAGCTGAGTAACAGCCAGCTTCTGGCTGACTTCTCTCCCCCAACGGTGGGA 841  
Qy 739 AACACAAATTTCTGTAAAGTACTACAGATGTAAATCTTGGGAAAATTTATAAATCTGT 798  
Db 842 GACACAGTTTCTAGATAAGTACTTCCAGATGTAAATCTTGGGAAAATTTATAAATCTGT 901  
Qy 799 TCTCGGAAAATTAATGAAGAGAAAGGTGAGATTTGGAACTTTTATCATGAATTTCAAT 858  
Db 902 TCTCGGAAAATTAATGAAGAGAAAGGTGAGATTTGGAACTTTTATCATGAATTTCAAT 961  
Qy 859 TAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCAAGTGAACCAATTTCTCAGGCC 918  
Db 962 TAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCAAGTGAACCAATTTCTCAGGCC 1021  
Qy 919 TACTTCAGAAAACAAAGAGAGCTTTTCAATGATCTGTTTAAATAATGAAGTGAACCGTGC 978  
Db 1022 CACTTCGGAACAAAGAGAGCTTTTCAAGCATCTGTTTAAAGTAAATGAAGTGAACCGTGC 1081  
Qy 979 TGAATAATACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGG 1038  
Db 1082 TGAGAACACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTGACGG 1141  
Qy 1039 AGTCTATGATTAATCTGATGTATGAGAGCGGTGATTTTCCAGGTTTCTGACTGGCTCA 1098  
Db 1142 AGTCTATGATTAATCTGATGTATGAGAGCGGTGATTTTCCAGGTTTCTGACTGGCTCA 1201  
Qy 1099 TCATCTCTTAATGGGAACTGGAATCTCTTTAAACCAACCTCGGAAAATGATGATGATTA 1158  
Db 1202 TCATCTCTTAATGGGAACTGGAATCTCTTTAAACCAACCTCGGAAAATGATGATGATTA 1261  
Qy 1159 CTATCTTCAGTGTAAACTAGAACAGCTATTTCCAGGAGACCGTTTGGTCTCACTCATAC 1218  
Db 1262 CTATCTTCAGTGTAAACTAGAACAGCTATTTCCAGGAGACCGTTTGGTCTCACTCATAC 1321  
Qy 1219 ACTTCTCAGAGATGCTATATTTCTGTAACCACTGAACTCGTCTCTCCAAAGATAAGCA 1278  
Db 1322 CCTTCTCAGAGATGCTATATTTTGTGAATACTGAACTCGTCTCTCCAAAGATAAGCA 1381  
Qy 1279 AAAAGGAGCAAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAA 1338  
Db 1382 AAAAGGAGCAAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAA 1441  
Qy 1339 GTGTATTTGGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCA 1398  
Db 1442 GTGTATTTGGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCA 1501  
Qy 1399 ACCAGTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATGATACAGAACTGTT 1458  
Db 1502 GCCAGTCTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATGATACAGAACTGTT 1561  
Qy 1459 TCCAGAGCTCAATAAGGTACAAAAGGAAGTTACTCTGTGACATCTTGGATGTAAACACT 1518  
Db 1562 TCCAGAGCTCAATAAGGTACAAAAGGAAGTTACTCTGTGACATCTTGGATGTAAACACT 1620  
Qy 1519 TGGATTTGGTATGAATAAACCAATTTGTTGTCGAGGAGTGTGAGAAATTTAC 1578  
Db 1621 TGGATTTGGTATGAATAAACCAATTTGTTGTCGAGGAGTGTGAGAAATTTAC 1658  
Qy 1579 TTTT-TTGGTATATCTTAT 1636  
Db 1659 TTTTCACTTTGTTATTTTATAAATATATATATATATATATATATATATATATATATAT 1712  
Qy 1637 TTTTGTTTTATAAAGACTTAACACAA 1664  
Db 1713 TTTTGTTTTATAAAGACTTAACAAATA 1740

## RESULT 11

LOCUS HSA420561 968 bp mRNA linear PRI 23-NOV-2001  
DEFINITION Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 1521079.  
ACCESSION AJ420561  
VERSION AJ420561.1 GI:17066425

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

1  
Auffray, C., Anorge, W., Ballabio, A., Estivill, X., Gibson, K.,  
Lehrach, H., Poustka, A. and Lundeberg, J.  
The European IMAGE consortium for integrated Molecular analysis of  
human gene transcripts

## JOURNAL

## REFERENCE

2 (bases 1 to 968)

## AUTHORS

Persson, A.

## TITLE

## JOURNAL

Direct Submission

Submitted (02-OCT-2001) Persson A., Center for Molecular  
Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagssvagen  
30B, 106 91 Stockholm, SWEDEN

## COMMENT

This clone is available royalty-free through IMAGE Consortium  
Distributors. IMPORTANT: This sequence represents the full insert  
of this IMAGE cDNA clone. No attempt has been made to verify  
whether this corresponds to the full-length of the original mRNA  
from which it was derived.

## FEATURES

source

1..968  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone\_lib="NCI CGAP Lu5"  
311 a 147 c 186 g - 324 t

## Query Match

46.5%; Score 925.4; DB 9; Length 968;

Best Local Similarity 99.8%; Pred. No. 1.3e-153;

Matches 937; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1019 GAGGTGATGACTGTAGAGGAGCTATGATTAATGGAATCGAATCTCTTAAAGAAACACC 1138  
Db 18 GGGGTGATGACTGTAGAGGAGCTATGATTAATGGAATCGAATCTCTTAAAGAAACACC 77  
Qy 1079 CAGGTTCCTGACTGGCTTCATCTCTTAATGGAATCGAATCTCTTAAAGAAACACC 1138  
Db 78 CAGGTTCCTGACTGGCTTCATCTCTTAATGGAATCGAATCTCTTAAAGAAACACC 137  
Qy 1139 CTGGAATGTATATCTGATTAATCTCTTCAAGTGAATCTCTTAAAGAAACACC 1198  
Db 138 CTGGAATGTATATCTGATTAATCTCTTCAAGTGAATCTCTTAAAGAAACACC 197  
Qy 1199 CGTTTGGTCTCACTCAATAACACTTCTCAGAGATGCTATATCTGTGAAAACACTGGAACCT 1258  
Db 198 CGTTTGGTCTCACTCAATAACACTTCTCAGAGATGCTATATCTGTGAAAACACTGGAACCT 257  
Qy 1259 CGCTCTCTCCAAAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATAC 1318  
Db 258 CGCTCTCTCCAAAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATAC 317  
Qy 1319 ATTCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTT 1378  
Db 318 ATTCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTT 377  
Qy 1379 CTGTTTGTATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTTGAC 1438  
Db 378 CTGTTTGTATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTTGAC 437  
Qy 1439 ATTGTGTATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGGAGTACCTCTGTG 1498  
Db 438 ATTGTGTATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGGAGTACCTCTGTG 497  
Qy 1499 ACATCTTGGATGAACACTTTGGATTTGGTATAGATAAACCATTGAAATTTTCGCTGTG 1558  
Db 498 ACATCTTGGATGAACACTTTGGATTTGGTATAGATAAACCATTGAAATTTTCGCTGTG 557

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-321N4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP1-3J17 is at 128911 in this sequence. The true right end of clone RP11-30P6 is at 100 in this sequence.

Location/Qualifiers

1. 129010  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-321N4"  
/clone\_lib="RPCI-11.2"  
2. 192  
/notes="MER21B repeat: matches 3. .207 of consensus"  
240. .909  
/notes="MER21B repeat: matches 30. .715 of consensus"  
1593. 1682  
/note="MIR repeat: matches 35. .145 of consensus"  
2030. .2102  
/note="L2 repeat: matches 2681. .2750 of consensus"  
3459. 3574  
/note="FLAM\_C repeat: matches 1. .117 of consensus"  
5635. 5670  
/note="L1ME repeat: matches 593. .718 of consensus"  
6887. 6982  
/note="L2 repeat: matches 2616. .2704 of consensus"  
8284. 8356  
/note="L2 repeat: matches 2639. .2710 of consensus"  
8380. 8405  
/note="L3 copies 2 mer tt 100% conserved"  
8440. 8601  
/note="MIR repeat: matches 1. .169 of consensus"  
8825. 9360  
/note="L2 repeat: matches 1685. .2301 of consensus"  
9423. 9592  
/note="L2 repeat: matches 2560. .2727 of consensus"  
12088. 12175  
/note="L4 copies 2 mer at 73% conserved"  
12110. 12177  
/note="L7 copies 4 mer atat 79% conserved"  
12790. 12905  
/note="L2 repeat: matches 2628. .2750 of consensus"  
13672. 13923  
/note="L2 repeat: matches 2462. .2725 of consensus"  
14774. 15053  
/note="AluSq repeat: matches 1. .289 of consensus"  
15966. 16031  
/note="L6 copies 11 mer 77% conserved"  
17304. 17359  
/note="MIR repeat: matches 196. .251 of consensus"  
17554. 17682  
/note="MIR repeat: matches 108. .228 of consensus"  
18366. 18701  
/note="MER2 repeat: matches 1. .345 of consensus"  
18959. 19248  
/note="AluJo repeat: matches 1. .284 of consensus"  
20386. 20592  
/note="L2 repeat: matches 70. .288 of consensus"  
20609. 20778  
/note="L1TR28 repeat: matches 948. .1013 of consensus"  
20747. 21117  
/note="L1RI repeat: matches 295. .680 of consensus"  
21163. 21226  
/note="L1TR28 repeat: matches 297. .361 of consensus"  
21187. 21359  
/note="L1RI repeat: matches 16. .188 of consensus"  
22191. 22355  
/note="FRAM repeat: matches 4. .163 of consensus"  
22844. 22991  
/note="MIR repeat: matches 48. .202 of consensus"  
23155. 23190

/note="18 copies 2 mer ga 86% conserved"  
24390..24440  
repeat\_region  
/note="L2 repeat: matches 2657. .2708 of consensus"  
25082..25109  
repeat\_region  
/note="7 copies 4 mer aac 100% conserved"  
25438..25744  
repeat\_region  
/note="L2 repeat: matches 2427. .2744 of consensus"  
25751..25917  
repeat\_region  
/note="MIR repeat: matches 95. .262 of consensus"  
25964..26307  
repeat\_region  
/note="L2 repeat: matches 1986. .2366 of consensus"  
26526..26673  
repeat\_region  
/note="FLAM C repeat: matches 1. .142 of consensus"  
26682..26737  
repeat\_region  
/note="L4 copies 4 mer agga 76% conserved"  
28208..28394  
repeat\_region  
/note="MIR repeat: matches 12. .196 of consensus"  
29052..29250  
repeat\_region  
/note="MIR repeat: matches 48. .251 of consensus"  
30082..30170  
repeat\_region  
/note="MIR repeat: matches 50. .145 of consensus"  
30197..31090  
repeat\_region  
/note="MER11D repeat: matches 1. .897 of consensus"  
31921..32067  
repeat\_region  
/note="FRAM repeat: matches -2. .151 of consensus"  
32545..32678  
repeat\_region  
/note="L2 repeat: matches 2578. .2708 of consensus"  
32860..32895  
repeat\_region  
/note="9 copies 4 mer ttta 83% conserved"  
32897..33193  
repeat\_region  
/note="AluX repeat: matches 1. .296 of consensus"  
35445..35478  
repeat\_region  
/note="MIR repeat: matches 84. .118 of consensus"  
37442..37598  
repeat\_region  
/note="MER5B repeat: matches 1. .173 of consensus"  
38559..38860  
repeat\_region  
/note="MIR repeat: matches 15. .236 of consensus"  
40552..40911  
repeat\_region  
/note="THE1B repeat: matches 1. .364 of consensus"  
40912..42540  
repeat\_region  
/note="THE1B-INTERNAL repeat: matches 1. .1580 of consensus"  
42541..42899  
repeat\_region  
/note="THE1B repeat: matches 1. .364 of consensus"  
42900..43231  
repeat\_region  
/note="MLT1A1 repeat: matches 4. .349 of consensus"  
43407..43851  
repeat\_region  
/note="MLT1C repeat: matches 9. .466 of consensus"  
44641..45020  
repeat\_region  
/note="MSTA repeat: matches 1. .382 of consensus"  
45689..45988  
repeat\_region  
/note="AluSg repeat: matches 1. .302 of consensus"  
46220..46351  
repeat\_region  
/note="Tigger3b repeat: matches 567. .698 of consensus"  
46358..46609  
repeat\_region  
/note="AluJb repeat: matches 40. .290 of consensus"  
46623..46758  
repeat\_region  
/note="AluJo/FRAM repeat: matches 170. .301 of consensus"  
46759..47229  
repeat\_region  
/note="Tigger3b repeat: matches 2. .540 of consensus"  
47230..47442  
repeat\_region  
/note="MER44A repeat: matches 85. .327 of consensus"  
47715..48005  
repeat\_region  
/note="L1P10 repeat: matches 5864. .6153 of consensus"  
48577..48973  
repeat\_region  
/note="MLT1B repeat: matches 1. .390 of consensus"  
50633..50925  
repeat\_region  
/note="AluY repeat: matches 3. .295 of consensus"  
53828..55056  
repeat\_region  
/note="L1MC1 repeat: matches 5023. .6332 of consensus"  
55092..56081  
repeat\_region  
/note="L1M4 repeat: matches 3778. .4798 of consensus"  
56074..56651  
repeat\_region

/note="L1M4 repeat: matches 579. .1150 of consensus"  
56652..56952  
repeat\_region  
/note="AluX repeat: matches 1. .301 of consensus"  
56953..57277  
repeat\_region  
/note="L1M4 repeat: matches 253. .579 of consensus"  
57280..57562  
repeat\_region  
/note="L2 repeat: matches 1590. .1886 of consensus"  
57699..57766  
repeat\_region  
/note="L2 repeat: matches 2639. .2710 of consensus"  
59295..59599  
repeat\_region  
/note="AluX repeat: matches 1. .307 of consensus"  
59749..59811  
repeat\_region  
  
Query Match 25.9%; Score 516; DB 9; Length 129010;  
Best Local Similarity 96.4%; Pred. No. 9 3e-82;  
Matches 528; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
  
QY 1440 TTGTGATACAGGAAGCTGTTTCCAGAGCTCAATAGGTACAAAAGGAAGTTACTCTGTGA 1499  
DB 50352 TTGCTCTTAAATGAATTTATTTTACTTTTAGGTACAAAAGGAAGTTACCTCTGTGA 50293  
  
QY 1500 CATCTTGGATGTAACACACCTGGATTTGGTATAGATAACCCATTGAAATTTCTGCTGTC 1559  
DB 50292 CATCTTGGATGTAACACCTGGATTTGGTATAGATAACCCATTGAAATTTCTGCTGTC 50233  
  
QY 1560 GAGGCTGGTAAATTTACTTTTTTGGGTATATTTCTTATATATATATATGATACATCGCTGT 1619  
DB 50232 GAGGCTGGTAAATTTACTTTTTTGGGTATATTTCTTATATATATATGATACATCGCTGT 50173  
  
QY 1620 CTGAAATTTTAGTTATTTTGTGTTTAAATAAGACTAACACAACTTAATGATTAAAG 1679  
DB 50172 CTGAAATTTTAGTTATTTTGTGTTTAAATAAGACTAACACAACTTAATGATTAAAG 50113  
  
QY 1680 TGATGTAGTCTCAGTCTTCTCAATTCCTAGCTGTGATGATCAATGATGTTATGCGCAACAGATA 1739  
DB 50112 TGATGTAGTCTCAGTCTTCTCAATTCCTAGCTGTGATGATCAATGATGTTATGCGCAACAGATA 50053  
  
QY 1740 TCATCTGTTATGGCATTTTTAAAGAGAAATTCATAATGATGTTATGCGCAACAGATA 1799  
DB 50052 TCATCTGTTATGGCATTTTTAAAGAGAAATTCATAATGATGTTATGCGCAACAGATA 49993  
  
QY 1800 AGACTGTAAACTTCGTATTGTATAGCTTTGAAATAATTTATGCTAGTATGAGAAACA 1859  
DB 49992 AGACTGTAAACTTCGTATTGTATAGCTTTGAAATAATTTATGCTAGTATGAGAAACA 49933  
  
QY 1860 GGAATAAGATCTGATTTTCTTAGAGTTAATATATTTTAGTAGATGTTTCCCTTTT 1919  
DB 49932 GGAATAAGATCTGATTTTCTTAGAGTTAATATATTTTAGTAGATGTTTCCCTTTT 49873  
  
QY 1920 TATTTGTACATAGTTAACTGTGTATCTATTAATAAAGCATCCTATATAGTTTATAA 1979  
DB 49872 TATTTGTACATAGTTAACTGTGTATCTATTAATAAAGCATCCTATATAGTTTATAA 49813  
  
QY 1980 ATAAAMA 1987  
DB 49812 ATAAACA 49805

RESULT 13  
AX331384/c  
LOCUS AX331384 451 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 1893 from Patent WO0194629.  
ACCESSION AX331384  
VERSION AX331384.1 GI:18122018  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
  
REFERENCE  
1  
AUTHORS Young P.E., Augustus M., Carter K.C., Ebner R., Endress G.,  
Horrigan S., Soppet D.R. and Weaver Z.  
TITLE Cancer gene determination and therapeutic screening using signature

Gene sets  
JOURNAL Patent: WO 0194629-A 1893 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES Location/Qualifiers  
source 1. 451  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 179 a 73 c 52 g 146 t 1 others  
ORIGIN  
Query Match 21.4%; Score 427.2; DB 6; Length 451;  
Best Local Similarity 98.9%; Pred. No. 1.4e-65;  
Matches 440; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1538 CCATTGAAATTTCTGCTGCGA-GGGTGTAGAAATTTACTTTTGGGTATATCTT 1596  
DB 445 CCATTGAAATTTCCGGCTGCGAGGGGTGTAGAAATTTACTTTTGGGTATATCTN 386  
QY 1597 ATATATATATGTACATCGCTGTCTGAAATTTTATTTTGTATTTTAAATAGACT 1656  
DB 385 ATATATATATGTACATCGCTGTCTGAAATTTTATTTTGTATTTTAAATAGACT 326  
QY 1657 AACCAAACTTAATGATTAAAGTATGAGTCTCATAGTCTTTTCATTTGCTAGCTGTGA 1716  
DB 325 AACCAAACTTAATGATTAAAGTATGAGTCTCATAGTCTTTTCATTTGCTAGCTGTGA 266  
QY 1717 TCCAAATTTTATGAAATTAAGTCACTTGTATTTGCCATTTTAAAGAGAAATTCAT 1776  
DB 265 TCCAAATTTTATGAAATTAAGTCACTTGTATTTGCCATTTTAAAGAGAAATTCAT 206  
QY 1777 AATGATGTTTGGCAACAGATAGATGATAAATTCGATTTGATATAGCTTTGAAATA 1836  
DB 205 AATGATGTTTGGCAACAGATAGATGATAAATTCGATTTGATATAGCTTTGAAATA 146  
QY 1837 ATTATGCTAGTATGAGAAACAGATAAGATCTGATTTTCTTATAGATTTAATATATTT 1896  
DB 145 ATTATGCTAGTATGAGAAACAGATAAGATCTGATTTTCTTATAGATTTAATATATTT 86  
QY 1897 AGTAGATGTTTTCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1956  
DB 85 AGTAGATGTTTTCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 26  
QY 1957 GCATCTATATGAGTTTATTAAT 1981  
DB 25 GCATCTATATGAGTTTATTAAT 1  
RESULT 14  
BD113855  
LOCUS 425 bp DNA linear PAT 18-SEP-2002  
DEFINITION EST and encoded human protein.  
ACCESSION BD113855  
VERSION BD13855.1 GI:3208759  
KEYWORDS JP 2002010789-A/5932.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 425)  
AUTHORS Edwards, J.B.D.M., Jobert, S. and Giordano, J.E.  
TITLE EST and encoded human protein  
JOURNAL Patent: JP 2002010789-A 5932 15-JAN-2002;  
GENSET CORP  
COMMENT OS Homo sapiens (human)  
PN JP 2002010789-A/5932  
PD 15-JAN-2002  
PF 07-AUG-2000 JP 2000280989  
PR 05-AUG-1999 US 60/147499  
PI JEAN BAPTIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI  
GIORDANO  
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N1/21,  
Gene sets  
JOURNAL Patent: WO 0194629-A 1893 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES Location/Qualifiers  
source 1. 425  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 134 a 82 c 85 g 124 t  
ORIGIN  
Query Match 21.3%; Score 425; DB 6; Length 425;  
Best Local Similarity 100.0%; Pred. No. 3.5e-65;  
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1004 CAGATATTTTATGGAGTGTATGATCTAGAGAGTCTATGATTACCTGATGATGTA 1063  
DB 1 CAGATATTTTATGGAGTGTATGATCTAGAGAGTCTATGATTACCTGATGATGTA 60  
QY 1064 GGACGGGTAGTTTCCAGGTTCCCTGACCTGCTTCAATCTCTTAAATGGGAACCTGATC 1123  
DB 61 GGACGGGTAGTTTCCAGGTTCCCTGACCTGCTTCAATCTCTTAAATGGGAACCTGATC 120  
QY 1124 CTCCTTAAACACCTCGAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1183  
DB 121 CTCCTTAAACACCTCGAATATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
QY 1184 CTATTTCCAGGAGCAGCTGTTGGTCTCTCACTCATACACATCTCTCAGAGATGCTATATCTGT 1243  
DB 181 CTATTTCCAGGAGCAGCTGTTGGTCTCTCACTCATACACATCTCTCAGAGATGCTATATCTGT 240  
QY 1244 GAAACACCTGAACTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGA 1303  
DB 241 GAAACACCTGAACTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGA 300  
QY 1304 GAAATGATGAATTAATTCAGATCTGATGATGATGATGATGATGATGATGATGATGAT 1363  
DB 301 GAAATGATGAATTAATTCAGATCTGATGATGATGATGATGATGATGATGATGATGAT 360  
QY 1364 GAAACATCAGACTCTCTGTTGATGGCTTACAGCAACAGTACTCAACAGCAGCTGACT 1423  
DB 361 GAAATGATGAATTAATTCAGATCTGATGATGATGATGATGATGATGATGATGATGAT 420  
QY 1424 TATGT 1428  
DB 421 TATGT 425  
RESULT 15  
BD026500  
LOCUS 358 bp DNA linear PAT 27-AUG-2002  
DEFINITION Sequence tag and encoded human protein.  
ACCESSION BD026500  
VERSION BD026500.1 GI:22567723  
KEYWORDS JP 2001269182-A/2746.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 358)  
AUTHORS Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.  
TITLE Sequence tag and encoded human protein  
JOURNAL Patent: JP 2001269182-A 2746 02-OCT-2001;  
GENSET  
COMMENT OS Homo sapiens (human)  
PN JP 2001269182-A/2746  
PD 02-OCT-2001  
PF 24-FEB-2000 JP 2000118773  
PR 26-FEB-1999 US 60/122487  
PI JEAN BAPTIST DUMAS MILNE EDWARDS, EIMERIC DUCCLAIR, JEAN YVES

PI JORDAN  
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10,  
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC  
G06F15/40

CC  
FH Key Location/Qualifiers  
FT CDS Location/Qualifiers  
1. 358  
53. 358.

FEATURES  
source  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 125 a 63 c 69 g 101 t

ORIGIN

Query Match 17.9%; Score 357; DB 6; Length 358;  
Best Local Similarity 100.0%; Pred. No. 3.7e-53;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 760 ACTACAGATGTAATCTTGGGAATTTATAAATCTGTCTCTGGAAACTAATGAAAGA 819  
DB 1 ACTACAGATGTAATCTTGGGAATTTATAAATCTGTCTCTGGAAACTAATGAAAGA 60  
QY 820 GAAAGTCTAGCATTTGGAACCTTTTATCATGAATTTTCAATTTCTTGAGTCTCCAAA 879  
DB 61 GAAAGTCTAGCATTTGGAACCTTTTATCATGAATTTTCAATTTCTTGAGTCTCCAAA 120  
QY 880 GCCTAAACCAAGTAGACCACTGACCACTTCTAGCCCTACTTTCAGAAACCAACAGAA 939  
DB 121 GCCTAAACCAAGTAGACCACTGACCACTTCTAGCCCTACTTTCAGAAACCAACAGAA 180  
QY 940 GCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAATACAGAGAAAGCA 999  
DB 181 GCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAATACAGAGAAAGCA 240  
QY 1000 AATATCAGATTTATTTATGAGGTGATGATCTAGAGGAGTCTATGATTACCTGATGA 1059  
DB 241 AATATCAGATTTATTTATGAGGTGATGATCTAGAGGAGTCTATGATTACCTGATGA 300  
QY 1060 TGTAGGACGGGTAGTTTCCAGGTTCTGAGTCTTCCATCTCTTAATGGGAC 1116  
DB 301 TGTAGGACGGGTAGTTTCCAGGTTCTGAGTCTTCCATCTCTTAATGGGAC 357

Search completed: January 31, 2004, 12:44:49  
Job time : 7419 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 10:23:38 : Search time 125 Seconds  
(without alignments)  
7033.882 Million cell updates/sec

Title: US-09-744-313A-3

Perfect score: 1992

Sequence: 1 gataaactcaaaagt.....tttaataataaaaaaaa 1992

Scoring table: IDENTITY NUC

Gapex 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA \*

- 1: /cgn2\_6/prodata/2/ina/5A COMB.seq \*
- 2: /cgn2\_6/prodata/2/ina/5B COMB.seq \*
- 3: /cgn2\_6/prodata/2/ina/6A COMB.seq \*
- 4: /cgn2\_6/prodata/2/ina/6B COMB.seq \*
- 5: /cgn2\_6/prodata/2/ina/PCUTS COMB.seq \*
- 6: /cgn2\_6/prodata/2/ina/backfiles1.seq \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1832.8	76.9	1716	4	US-09-620-312D-922
2	1392.8	59.9	1551	4	US-09-620-312D-923
3	57.2	2.9	658	3	US-08-998-416-595
C 4	53.8	2.7	7218	1	US-08-232-463-14
C 5	53.2	2.7	8920	2	US-08-446-855A-1
C 6	53.2	2.7	8920	3	US-09-150-741-1
C 7	51.2	2.6	6124	4	US-08-213-419B-3
C 8	48.2	2.4	2110	3	US-09-419-459-1
C 9	47.4	2.4	615	3	US-08-998-416-186
C 10	47.4	2.4	665	2	US-08-883-795A-36
C 11	44.6	2.2	837	3	US-08-998-416-288
C 12	44.4	2.2	6152	3	US-08-973-462-1
C 13	44	2.2	1851	4	US-09-601-198-51
C 14	43.8	2.2	483	4	US-09-220-132-18
C 15	43.8	2.2	483	4	US-09-220-132-161
C 16	43.8	2.2	662	3	US-08-998-416-185
C 17	43.8	2.2	701	3	US-08-998-416-701
C 18	43.8	2.2	724	3	US-08-998-416-683
C 19	43.8	2.2	732	3	US-08-998-416-1036
C 20	43.8	2.2	767	3	US-08-998-416-472
C 21	43.8	2.2	827	3	US-08-998-416-535
C 22	43.8	2.2	828	3	US-08-998-416-538
C 23	43.8	2.2	834	3	US-08-998-416-305
C 24	43.8	2.2	10640	4	US-09-417-485D-5
C 25	43.6	2.2	19124	2	US-08-487-826B-13
C 26	43.4	2.2	636	3	US-08-998-416-1137
C 27	43.4	2.2	765	4	US-09-601-198-160

C 28	43.4	2.2	1430	2	US-08-179-557-15	Sequence 15, Appl
C 29	43.2	2.2	1887	4	US-09-601-198-39	Sequence 39, Appl
C 30	43	2.2	1186	2	US-08-731-722-5	Sequence 5, Appl
C 31	42.8	2.1	148567	4	US-09-801-876B-3	Sequence 3, Appl
C 32	42.6	2.1	711	3	US-08-998-416-786	Sequence 786, App
C 33	42.4	2.1	1230025	4	US-09-198-452A-1	Sequence 1, Appl
C 34	42.2	2.1	665	3	US-08-998-416-937	Sequence 7, Appl
C 35	42.2	2.1	2030	2	US-08-705-937-7	Sequence 1, Appl
C 36	42.2	2.1	168575	4	US-09-426-290-1	Sequence 1, Appl
C 37	42	2.1	1497	4	US-09-620-312D-1021	Sequence 1021, Ap
C 38	42	2.1	6124	4	US-08-213-419B-3	Sequence 3, Appl
C 39	41.6	2.1	8920	2	US-08-446-855A-1	Sequence 1, Appl
C 40	41.6	2.1	8920	3	US-09-150-741-1	Sequence 1, Appl
C 41	41.6	2.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 42	41.4	2.1	417	4	US-08-559-896B-3	Sequence 3, Appl
C 43	41.4	2.1	639	4	US-09-328-352-1706	Sequence 1706, Ap
C 44	41.4	2.1	1956	4	US-08-559-896B-1	Sequence 1, Appl
C 45	41.2	2.1	3060	4	US-09-996-243-504	Sequence 504, App

ALIGNMENTS

RESULT 1

US-09-620-312D-922

; Sequence 922, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Mehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yuning

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. 6569662el Nucleic Acids and

; FILE REFERENCE: 784CIP2B

; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: Pf FL\_genes Version 1.0

; SEQ ID NO 922

; LENGTH: 1716

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (71)..(1543)

US-09-620-312D-922

Query Match 76.9%; Score 1532.8; DB 4; Length 1716;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 143 TTCAGGACACACAGAAAACGGGAGATCATTTGGATCAGCAGATAGTAGCAAAATT 202

Db 173 TTTCGGAACACACAGAAAACGGGAGATCATTTGGATCAGCAGATAGTAGCAAAATT 232



203 AAAGGATATTCAGAAAGTACCAATGAGGAGCTATGTTGCTTAATATGTTAGCT 262  
233 AAAGGATATTCAGAAAGTACCAATGAGGAGCTATGTTGCTTAATATGTTAGCT 292  
263 GAAGTGAAGATGATTTTATGAAAGAGTATGTTGCTTAATGAAAGATGATTTCCAGTG 322  
293 GAAGTGAAGATGATTTTATGAAAGAGTATGTTGCTTAATGAAAGATGATTTCCAGTG 352  
323 GAAGTGAAGATGATTTTATGAAAGAGTATGTTGCTTAATGAAAGATGATTTCCAGTG 382  
353 GAAGTGAAGATGATTTTATGAAAGAGTATGTTGCTTAATGAAAGATGATTTCCAGTG 412  
383 TATGTAGATTTTATGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 442  
413 TATGTAGATTTTATGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 472  
443 GTGTTTGTATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 502  
473 GTGTTTGTATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 532  
503 TGGTCTGTATGAAAGATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 562  
533 TGGTCTGTATGAAAGATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 592  
563 CATGTCATTTCTGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 622  
593 CATGTCATTTCTGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 652  
623 GAATTTCTTAAAGTCAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 682  
653 GAATTTCTTAAAGTCAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 712  
683 CCAGAACTGAGTATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 742  
713 CCAGAACTGAGTATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 772  
743 CAATTTCTGATGAAAGATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 802  
773 CAATTTCTGATGAAAGATGATTTGAAAGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 832  
803 GGAAACTTAAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 862  
833 GGAAACTTAAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 892  
863 TCTGTGAGTCTCCAAAGCTTAAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 922  
893 TCTGTGAGTCTCCAAAGCTTAAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 952  
923 TCAGAAACCAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 982  
953 TCAGAAACCAAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1012  
983 AATCAGAGAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1042  
1013 AATCAGAGAGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1072  
1043 TATGATTAACCTGATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1102  
1073 TATGATTAACCTGATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1132  
1103 CTCTTAATGGAAGTATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1162  
1133 CTCTTAATGGAAGTATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1192  
1163 CTCTTAATGGAAGTATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1222  
1193 CTCTTAATGGAAGTATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1252  
1223 CTCTTAATGGAAGTATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1282  
1253 CTCTTAATGGAAGTATGATGAGGAGTATGTTGCTTAATGAAAGAGTATGATTTCCCT 1312  
1283 GGAGCAAAACAGACTTTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1342

1313 GGAGCAAAACAGACTTTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1372  
1343 ATTGTGAGAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1402  
1373 ATTGTGAGAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1432  
1403 GTACTCAACAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1462  
1433 GTACTCAACAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1492  
1463 GAGCTCAATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1522  
1493 GAGCTCAATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1552  
1523 TTTGCTATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1582  
1553 TTTGCTATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1612  
1583 TTTGCTATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1642  
1613 TTTGCTATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1672  
1643 TTTTAAATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1678  
1673 TTTTAAATAGAGTATGATTTGAAGAGTATGATTTCCCTTAATGATTTCCCTTAATGATTTCCCT 1708

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1378)  
US-09-620-312D-923

Query Match 59.9%; Score 1192.8; DB 4; Length 1551;  
Best Local Similarity 89.1%; Pred. NO. 6.8e-269;  
Matches 1369; Conservative 0; Mismatches 2; Indels 165; Gaps 1;

143 TTCAAGAACACACAGAAAGGGGAGGATCATTTGGATCAGCAGAAATAGTAGCAAAAT 202  
Db  
173 TTTCGGAACACACAGAAAGGGGAGGATCATTTGGATCAGCAGAAATAGTAGCAAAAT 232  
Qy 203 AAAGAGTATTCAAAAGTACCAACAATGGAGGAGCTATGTTGCTTAATATGCTAGCT 262  
Db 233 AAAGAGTATTCAAAAGTACCAACAATGGAGGAGCTATGTTGCTTAATATGCTAGCT 292  
Qy 263 AAAGAGTATTCAAAAGTATTCAAAAGTATTCAAAAGTATTCAAAAGTATTCAAAAGT 322  
Db 293 AAAGAGTATTCAAAAGTATTCAAAAGTATTCAAAAGTATTCAAAAGTATTCAAAAGT 352  
Qy 323 GAGCTGTGAGCAACCTTACTCCCGAAACCTTCTGTCATGGAATATGCAATTCAC 382  
Db 353 GAGCTGTGAGCAACCTTACTCCCGAAACCTTCTGTCATGGAATATGCAATTCAC 412  
Qy 383 TATGTAGACTTTTGGAGTCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 442  
Db 413 TATGTAGACTTTTGGAGTCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 472  
Qy 443 GTGTTTGTATGATGTTGAAAGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 502  
Db 473 GTGTTTGTATGATGTTGAAAGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532  
Qy 503 TGTCTGTCTATAGAGATATCTGAAATCTTGAATCTGAAATCTGAAATCTGAAATCTGAA 562  
Db 533 TGTCTGTCTATAGAGATATCTGAAATCTTGAATCTGAAATCTGAAATCTGAAATCTGAA 592  
Qy 563 CATGCTGCTATCTGAGTCCAGCTTCTTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAG 622  
Db 593 CATGCTGCTATCTGAGTCCAGCTTCTTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAG 652  
Qy 623 GAATTTTAAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 682  
Db 653 GAATTTTAAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 712  
Qy 683 CCAGAGTGTATGATCAACTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 742  
Db 713 CCAGAGTGTATGATCAACTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 772  
Qy 743 CAATTTCTGTATGATCAACTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 802  
Db 773 CAATTTCTGTATGATCAACTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 832  
Qy 803 GGAAAGTCAATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 862  
Db 833 GGAAAGTCAATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 892  
Qy 863 TCTTGTGAGTCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 922  
Db 893 TCTTGTGAGTCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 952  
Qy 923 TCAGAAACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982  
Db 953 TCAGAAACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1012  
Qy 983 AATACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1042  
Db 1013 AATACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1072  
Qy 1043 TATGATTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
Db 1073 TATGATTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1092  
Qy 1103 CTCTTAATGGGAACCTGAAATCTCTTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1162  
Db 1093 ----- 1092  
Qy 1163 CTTTCACTGTAACTAGACAGCTATTTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1222  
Db 1093 ----- 1092

1223 CTCAGAGATGCTATATTTCTGTGAAACACCTGCTCTCTCCAGATAGCAAAAAA 1282  
Db 1093 -----AGATGCTATATTTCTGTGAAACACCTGCTCTCTCCAGATAGCAAAAAA 1147  
Qy 1283 CGAGCAAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1342  
Db 1148 CGAGCAAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1207  
Qy 1343 ATTGGTGAAGAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGAT 1402  
Db 1208 ATTGGTGAAGAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGAT 1267  
Qy 1403 GTACTCAACAGAGCTGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGAT 1462  
Db 1268 GTACTCAACAGAGCTGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGAT 1327  
Qy 1463 GAGCTCAATAGAGTACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1522  
Db 1328 GAGCTCAATAGAGTACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1387  
Qy 1523 TTTGGTATAGATTAACCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1582  
Db 1388 TTTGGTATAGATTAACCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447  
Qy 1583 TTTGGTATATTTCTTAT 1642  
Db 1448 TTTGGTATATTTCTTAT 1507  
Qy 1643 TTTTAAATAAGACTAACACAACTTAATGATTAATAA 1678  
Db 1508 TTTTAAATAAGACTAACACAACTTAATGATTAATAA 1543

## RESULT 3

US-08-998-416-595  
; Sequence 595, Application US/08998416  
; Patent No 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippen, Peter  
; APPLICANT: Fohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jergen  
; APPLICANT: Knechtel, Philipp  
; APPLICANT: Reibschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587



```

RESULT 6
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02

```

```

RESULT 7
US-08-213-419B-3/c
; Sequence 3, Application US/082134.9B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: JII-002CNCF
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

```



1752 GCATTTTAAAGAGAAATTCATAATGATGTTATGGCAACAGATAGACTGATAAC 1811  
372 AATATTTGTAATTTATTTTATTAATAATCTATTTTATAAATTTATGTTGTTT 431  
1812 TTGCTATTGTAAGCTTTGAAATAATTAATGCTAGTAGGAGAAACAGGAATAGATCT 1871  
432 ATATATTTAACTTTTAAAGAAATTAATTTAAATTTAACTTTTAACTTTCTTAT 491  
1872 GATTTCTTAGAGTTAAATATTTTAGTAGATTTGTTTCTTTTATTTTATTTGTCAT 1931  
492 TATTAATTTTATATTTATTTAATAATTTATATTTTATTTTATTTTATTTAATTA 551  
1932 AGTTAACTGCTATCTATATAATAAAGACATCCATATGATGTTTATTAATAAATAA 1991  
552 AATTAATTTATTTAATTTAATTTTATTTTATTTTATTTTATTTAATTAATAAATA 611  
1992 A 1992  
612 A 612

RESULT 10  
US-08-883-795A-36/c  
Sequence 36, Application US/08883795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcove, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERSKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 665 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE: ORGANISM: Homo sapiens  
IMMEDIATE SOURCE: CLONE: Rh 32  
US-08-883-795A-36  
Query Match 2.4% Score 47.4; DB 2; Length 665;  
Best Local Similarity 47.2%; Pred. No. 0.051;  
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;  
1628 TTACTATTTTGTGTTTTTAAAGACTAACCAACTTAATGATTAAGTATGATTTGAG 1687

344 TTATAACATTTTAAATTAATAATATGTAATTAATAACATTTTAAATTAATAATTTTAA 285  
1688 TCTCAGTAGCTTTTCATTTGCTAGCTGATCCAAATTTTATTAGACATAAGTCACTTGT 1747  
284 TTATAACATTTTAAATTAATAATATTTAAATTAATAATTTTAAATTAATAATTTTAA 225  
1748 TATTGCCATTTTAAAGAGAAATTCATAATGATGTTATGGCAACAGATAGACTGAT 1807  
224 TTATAAATTTTAAATTAATAATATTTAAATTAATAATTTTAAATTAATAATTTTAA 165  
1808 AACTTCGATTTGTAAGCTTTGAAATAATTAATGCTAGTAGTGGAGAAACAGGAATAAG 1867  
164 TTATAAATTTTAAATTAATAATATTTAAATTAATAATTTTAAATTAATAATTTTAA 105  
1868 ATCTGATTTTCTAGAGTTAATAATTTTAGTAGATTTGTTTCTTTTATTTTATTTTGT 1927  
104 TTATAAATTTTAAATTAATAATATTTAAATTAATAATTTTAAATTAATAATTTTAA 45  
1928 ACATA 1932  
44 TTATA 40

RESULT 11  
US-08-998-416-288  
Sequence 288, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgin  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: ORGANISM: PAG1241RP





GENERAL INFORMATION:  
APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 07334-074001  
CURRENT APPLICATION NUMBER: US/09/220,132  
CURRENT FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: US 60/079,303  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: US 60/068,821  
PRIOR FILING DATE: 1997-12-24  
NUMBER OF SEQ ID NOS: 191  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 483  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(483)  
OTHER INFORMATION: n = A, T, C or G  
US-09-220-132-18

Query Match 2.2%; Score 43.8; DB 4; Length 483;  
Best Local Similarity 54.4%; Pred. No. 0.31;  
Matches 87; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 1822 ATAGCTTTGAAATAAATTATGCTAGTAGTGGAGAAACAGGAATAAGATCTGATTTTCTTA 1881  
Db 392 ATGGGTTTAAATTAATAAGAGCATCCGGTTTGGTATGGGATGATCCAGGATTAATGTTG 333  
QY 1882 GAGTTAATAATTTTAGTAGATGGTTTCCCTTTTATTTTGTACATAGTAACTGT 1941  
Db 332 TGACTGATACATATTAGTTACTTGTCCTTTTTTTTTTTTTTTTGGATCTTTGCAAGGC 273  
QY 1942 GTATCTATAATAAAGCATCCCTATATGAGTTTATAATAAT 1981  
Db 272 AAACACAGTAACGAGTTTATATAATAATTAAATT 233

RESULT 15  
US-09-220-132-161/C  
Sequence 161, Application US/09220132  
Patent No. 6506607  
GENERAL INFORMATION:  
APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 07334-074001  
CURRENT APPLICATION NUMBER: US/09/220,132  
CURRENT FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: US 60/079,303  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: US 60/068,821  
PRIOR FILING DATE: 1997-12-24  
NUMBER OF SEQ ID NOS: 191  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 161  
LENGTH: 483  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(483)  
OTHER INFORMATION: n = A, T, C or G  
US-09-220-132-161

Query Match 2.2%; Score 43.8; DB 4; Length 483;  
Best Local Similarity 54.4%; Pred. No. 0.31;  
Matches 87; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 1822 ATAGCTTTGAAATAAATTATGCTAGTAGTGGAGAAACAGGAATAAGATCTGATTTTCTTA 1881  
Db 392 ATGGGTTTAAATTAATAAGAGCATCCGGTTTGGTATGGGATGATCCAGGATTAATGTTG 333  
QY 1882 GAGTTAATAATTTTAGTAGATGGTTTCCCTTTTATTTTGTACATAGTAACTGT 1941  
Db 332 TGACTGATACATATTAGTTACTTGTCCTTTTTTTTTTTTTTTTGGATCTTTGCAAGGC 273  
QY 1942 GTATCTATAATAAAGCATCCCTATATGAGTTTATAATAAT 1981  
Db 272 AAACACAGTAACGAGTTTATATAATAATTAAATT 233

Db 392 ATGGGTTTAAATTAATAAGAGCATCCGGTTTGGTATGGGATGATCCAGGATTAATGTTG 333  
QY 1882 GAGTTAATAATTTTAGTAGATGGTTTCCCTTTTATTTTGTACATAGTAACTGT 1941  
Db 332 TGACTGATACATATTAGTTACTTGTCCTTTTTTTTTTTTTTTTGGATCTTTGCAAGGC 273  
QY 1942 GTATCTATAATAAAGCATCCCTATATGAGTTTATAATAAT 1981  
Db 272 AAACACAGTAACGAGTTTATATAATAATTAAATT 233

Search completed: January 31, 2004, 13:58:47  
Job time : 130 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 09:03:22 ; Search time 547 Seconds

(without alignments)  
9830.495 Million cell updates/sec

Title: US-09-744-313A-3

Perfect score: 1992

Sequence: 1-gtataaaactccaaagt.....tttaataataaaaaaaaa 1992

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756-seq, 134219017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1992	100.0	1992	22 AAC86397	SNEXN DNA #1. Hom
C 2	1840.2	92.4	2176	22 AAI60815	Human polynucleoti
C 3	1840.2	92.4	2176	22 AAI60816	Human polynucleoti
4	1532.8	76.9	1715	22 AAI59029	Human polynucleoti
5	1504	75.5	3145	24 ABQ79518	CDNA encoding a pr
6	1192.8	59.9	1551	22 AAI59030	Human polynucleoti
7	655.4	32.9	779	20 AAZ17379	Human gene express
C 8	427.2	21.4	451	24 ABL63556	Breast cancer rela

9	424.4	21.3	725	24	ABS77486	Frog embryonic gen
10	364.2	18.3	402	25	ABX50611	Bovine EST associa
11	357	17.9	358	21	AAC02755	Human secreted pro
12	267.2	13.4	305	25	ABX50426	Bovine EST associa
13	86	4.3	311	22	AAH69552	Human cervical can
14	86	4.3	313	22	AAH70887	Human cervical can
15	84	4.2	299	22	AAH72445	Human cervical can
16	84	4.2	299	22	AAH73007	Human cervical can
C 17	78.6	3.9	580	24	ABN62287	Human cancer relat
18	68.4	3.4	9095	24	ABQ67061	Human angiogenesis
C 19	67.2	3.4	8056	25	ABZ10246	Haematopoietic cel
20	66.2	3.3	7461	24	ABL33785	Human immune syste
21	65.6	3.3	6161	24	ABL32623	Human immune syste
C 22	65.6	3.3	8056	25	ABZ10100	Haematopoietic cel
23	63.4	3.2	7306	24	ABN80291	Human chemically m
24	63.4	3.2	18133	24	ABK40017	Human chemically p
25	63.4	3.2	18133	24	ABL32940	Human immune syste
26	62.6	3.1	4590	7	AAH60472	Sequence encoding
27	62.2	3.1	73334	24	ABL92319	Chemically treated
28	62.2	3.1	73334	24	ABL34125	Human immune syste
29	62	3.1	14316	24	ABL70605	Chemically treated
30	62	3.1	14316	24	AAS61444	Human gene regulat
31	62	3.1	14316	24	ABK31518	Signal transductio
32	61.6	3.1	15767	24	ABL33207	Human immune syste
33	61.6	3.1	15767	24	ABL34553	Human metastasis a
34	61.6	3.1	17183	24	ABL32487	Human immune syste
35	61.4	3.1	5997	24	ABL33624	Human immune syste
36	60.8	3.1	8056	25	ABZ10246	Haematopoietic cel
37	60.6	3.0	5413	22	AAS46694	Tumour suppressor
38	60.2	3.0	61020	22	AAS46788	Tumour suppressor
39	60	3.0	60	60	ABN39970	Human spliced tran
40	60	3.0	6090	25	ABZ10150	Haematopoietic cel
41	59.8	3.0	6203	22	AAS45476	Chemically pretrea
42	59.8	3.0	6203	24	ABK28400	DNA transcription
C 43	59.4	3.0	612	22	AAH71471	Human cervical can
44	59.4	3.0	9145	24	ABL32888	Human immune syste
45	59.4	3.0	17389	24	ABL33415	Human immune syste

#### ALIGNMENTS

RESULT 1  
AAC86397  
ID AAC86397 standard; DNA; 1992 BP.  
XX  
AC AAC86397;  
XX  
DT 21-MAR-2001 (first entry)  
XX  
DE SNEXN DNA #1.  
XX  
KW SNEXN; human; sorting nexin; inflammation; asthma; allergy; AIDS;  
KW neurological disorder; gastrointestinal; smooth muscle cell;  
KW cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200073334-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 26-MAY-2000; 2000WO-US14831.  
XX  
PR 27-MAY-1999; 98US-0136740.  
PR 16-JUN-1999; 99US-0139566.  
PA (INCV-) INCYTE GENOMICS INC.  
XX  
PI Yue H, Tang YT, Arimazai Y;  
XX WPI; 2001-041141/05.  
XX

PT Novel Sorting Nexin polypeptides and polynucleotides useful for  
PT diagnosing and treating disorders associated with their expression e.g.  
PT autoimmune disorders, smooth muscle cell disorders and cell  
XX proliferative disorders  
PS Claim 5; Page 82-83; 84pp; English.  
XX  
CC The present invention relates to human sorting nexin (SNEXN).  
CC Compositions containing SNEXN or agonists of SNEXN are useful for  
CC treating a disease or condition associated with decreased expression of  
CC functional SNEXN and compositions containing antagonists of SNEXN are  
CC useful for treating a disease or condition associated with  
CC overexpression of functional SNEXN. These can be immune  
CC disorders for example inflammation, asthma, allergy, and AIDS,  
CC neurological disorders, gastrointestinal disorders, smooth muscle cell  
CC disorders, cancers and others. The SNEXN may also be used for somatic  
CC or germline gene therapy.  
XX  
SQ Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;

Query Match 100.0%; Score 1992; DB 22; Length 1992;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATGAAGAACTCCAAAGTTATGTAGCTATACCTTTAGTTTATCATTTTCAAAACTGTT 60  
DB 1 GTATGAAGAACTCCAAAGTTATGTAGCTATACCTTTAGTTTATCATTTTCAAAACTGTT 60

QY 61 TTTCTTTTATTAATGTTTTTCAATAACTAGACTGTGTGATAAATTTGCTATGTA 120  
DB 61 TTTCTTTTATTAATGTTTTTCAATAACTAGACTGTGTGATAAATTTGCTATGTA 120

QY 121 CTTGTATACATTTTGTGTTAATATTCAGGAACAACAGAAAGGGGAGGAATCATTTGGAAT 180  
DB 121 CTTGTATACATTTTGTGTTAATATTCAGGAACAACAGAAAGGGGAGGAATCATTTGGAAT 180

QY 181 CAGCAGATAGGTAGCAAAATTAAGAGGTATTTCAAAAGTACCACAAATGGAGGAGCTAT 240  
DB 181 CAGCAGATAGGTAGCAAAATTAAGAGGTATTTCAAAAGTACCACAAATGGAGGAGCTAT 240

QY 241 GTTGCTTAATTTGGTGTAGCTGAGGTGAAGTGAAGTATTTTGAAGAGGTATTTGTTGT 300  
DB 241 GTTGCTTAATTTGGTGTAGCTGAGGTGAAGTGAAGTATTTTGAAGAGGTATTTGTTGT 300

QY 301 AATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGC 360  
DB 301 AATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGC 360

QY 361 TGCATGGAATAATAGCAATCCATATGATAGACTTTTGTGAGGATCCCTCTCTGAAAGGAA 420  
DB 361 TGCATGGAATAATAGCAATCCATATGATAGACTTTTGTGAGGATCCCTCTCTGAAAGGAA 420

QY 421 GGGGAAAGAAAGAAATTCCTGTGTTTGTATGTTGTTGAAGAAATGATAGAGAGC 480  
DB 421 GGGGAAAGAAAGAAATTCCTGTGTTTGTATGTTGTTGAAGAAATGATAGAGAGC 480

QY 481 AGTTGACACGAGCCTGGAACATTTGCTCTATAGAGATATCTTGAATTTCTATGTA 540  
DB 481 AGTTGACACGAGCCTGGAACATTTGCTCTATAGAGATATCTTGAATTTCTATGTA 540

QY 541 TGAATCAAACTAACAGAAATTTTCATGCTGCTATGATGCTCCAGCTTCTTCTPAAGAG 600  
DB 541 TGAATCAAACTAACAGAAATTTTCATGCTGCTATGATGCTCCAGCTTCTTCTPAAGAG 600

QY 601 GATCATTGGCCCCCAAAATTAAGAAATTTTAAAGTCAAGAGGGAGAGTTCACAGATA 660  
DB 601 GATCATTGGCCCCCAAAATTAAGAAATTTTAAAGTCAAGAGGGAGAGTTCACAGATA 660

QY 661 TCTACAGAACTTCTGAGCATCCAGAACTGAGTAAATAGTCAACTTCTGGCAGACTTCT 720  
DB 661 TCTACAGAACTTCTGAGCATCCAGAACTGAGTAAATAGTCAACTTCTGGCAGACTTCT 720

QY 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGTATAGATACCTACAGATGTAATCTTGG 780

DB 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGTATAGATACCTACAGATGTAATCTTGG 780  
QY 781 GAAATTAATAAATCTGTTCTGGAAACCTAATTAAGAGAAAGGTACAGATTGGAACC 840  
DB 781 GAAATTAATAAATCTGTTCTGGAAACCTAATTAAGAGAAAGGTACAGATTGGAACC 840

QY 841 TTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGACCAGA 900  
DB 841 TTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGACCAGA 900

QY 901 ACTGACCATTTCTCAGCCCTTACTTTCAGAAAAACAACAGAGCTTTTCAATGATCTGTTAA 960  
DB 901 ACTGACCATTTCTCAGCCCTTACTTTCAGAAAAACAACAGAGCTTTTCAATGATCTGTTAA 960

QY 961 AATAATCAAAACCGTGTGTAATAATACAGAGAGAAAGCAAAATCAGAAATTTATGGA 1020  
DB 961 AATAATCAAAACCGTGTGTAATAATACAGAGAGAAAGCAAAATCAGAAATTTATGGA 1020

QY 1021 GGTGATGCTGTAGAGGAGTCTATGATTTACCTGTATGTAGGACGGTAGTTTCCA 1080  
DB 1021 GGTGATGCTGTAGAGGAGTCTATGATTTACCTGTATGTAGGACGGTAGTTTCCA 1080

QY 1081 GGTTCCTGACTGGCTTCATCTCTTAATGGAACTCGAATCCTCTTTAAAAACACCCCT 1140  
DB 1081 GGTTCCTGACTGGCTTCATCTCTTAATGGAACTCGAATCCTCTTTAAAAACACCCCT 1140

QY 1141 GGAATGTATATGATTTACTTCTCAGTGTAAACCTAGAAACAGCTATTTTCAGAGACCG 1200  
DB 1141 GGAATGTATATGATTTACTTCTCAGTGTAAACCTAGAAACAGCTATTTTCAGAGACCG 1200

QY 1201 TTTGGTCTCACTCATTAACACCTTCTCAGAGATGCTATATTTCTGTGAAAACTGAACTCG 1260  
DB 1201 TTTGGTCTCACTCATTAACACCTTCTCAGAGATGCTATATTTCTGTGAAAACTGAACTCG 1260

QY 1261 CTCTCTCCAGATAAAGCAAAAGGAGCAAAAACAGACTTTTGAAGAAATCATGAATTACAT 1320  
DB 1261 CTCTCTCCAGATAAAGCAAAAGGAGCAAAAACAGACTTTTGAAGAAATCATGAATTACAT 1320

QY 1321 TCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCT 1380  
DB 1321 TCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCT 1380

QY 1381 GTTGTATGGCTTACAGAAACCGAGTACTCAACAGAGCTGACTTATGTTTATTTGACAT 1440  
DB 1381 GTTGTATGGCTTACAGAAACCGAGTACTCAACAGAGCTGACTTATGTTTATTTGACAT 1440

QY 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAAAGTACCTCTGTGAC 1500  
DB 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAAAGTACCTCTGTGAC 1500

QY 1501 ATCTTGGATGTAACACCTTGGATTTGGTATAGAAATTAACCCATTTGAAATTTCTGCTGCG 1560  
DB 1501 ATCTTGGATGTAACACCTTGGATTTGGTATAGAAATTAACCCATTTGAAATTTCTGCTGCG 1560

QY 1561 AGGTGGTAGAAATTTACTTTTTTGGTATATTTCTTATATATTTATGATCATCGCTGTC 1620  
DB 1561 AGGTGGTAGAAATTTACTTTTTTGGTATATTTCTTATATATTTATGATCATCGCTGTC 1620

QY 1621 TGAATTTTATGTTTATTTTGTGTTTAAATAAGACTTAAACAACTTAATGATTAAGT 1680  
DB 1621 TGAATTTTATGTTTATTTTGTGTTTAAATAAGACTTAAACAACTTAATGATTAAGT 1680

QY 1681 GATTGAGTCTCATAGTCTTTCATTTCTGTAGCTGTGATCCAAATTTTATTTAGAACATAAGT 1740  
DB 1681 GATTGAGTCTCATAGTCTTTCATTTCTGTAGCTGTGATCCAAATTTTATTTAGAACATAAGT 1740

QY 1741 CACTTGTATTGCAATTTTAAAGAGAAATTCATAATGATGTTTATGGAACACAGATAA 1800  
DB 1741 CACTTGTATTGCCAATTTTAAAGAGAAATTCATAATGATGTTTATGGAACACAGATAA 1800

QY 1801 GACTGATAAATCTGTTATGTTATAGCTTTGAAATTAATTTATGCTAGTATGGAGAACAG 1860

Db 1801 GACTGATAAACTTGGTATTGTATAGCTTTGAAATAATATTGCTAGTATGGAGAAACAG 1860  
Qy 1861 GAATAAGACTGCTGATTTCTTAGAGTAAATATATTTTAGTAGATGTTTCCCTTTT 1920  
Db 1861 GAATPARGACTGATTTCTTAGAGTAAATATATTTTAGTAGATGTTTCCCTTTT 1920  
Qy 1921 ATTTTGACATAGTAACTGATGATCTATATAAATAAGCATCCTATATGAGTTTATAA 1980  
Db 1921 ATTTTGACATAGTAACTGATGATCTATATAAATAAGCATCCTATATGAGTTTATAA 1980  
Qy 1981 TAAAAA 1992  
Db 1981 TAAAAA 1992

RESULT 2  
AAI60815/c  
ID AAI60815 standard, cDNA; 2176 BP.  
AC AAI60815;  
XX  
DT 22-OCT-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 4804.  
DE  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX  
XX 25-APR-2000; 2000US-0552317.  
XX  
XX 09-JUL-2000; 2000US-0598042.  
XX  
XX 19-JUL-2000; 2000US-0620312.  
XX  
XX 03-AUG-2000; 2000US-0653450.  
XX  
XX 14-SEP-2000; 2000US-0662191.  
XX  
XX 19-OCT-2000; 2000US-0693036.  
XX  
XX 29-NOV-2000; 2000US-0727344.  
XX

(HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao Q, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
XX P-PSDB; AAM41659.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders.  
XX such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 4804; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
Query Match 92.4%; Score 1840.2; DB 22; Length 2176;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1842; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 143 TTCCAGGAAACACACAGAAAAGGGAGATCATTTTGGAAATCAGCAGATAGTAGCAAAATT 202  
Db 2022 TTTCGAAACACACAGAAAAGGGAGATCATTTTGGAAATCAGCAGATAGTAGCAAAATT 1963  
Qy 203 AAAGGAGTATTCAAAGTACCACCAATGAGGAGCTATTTGTCCTTAATTTAGGTAGCT 262  
Db 1962 AAAGGAGTATTCAAAGTACCACCAATGAGGAGCTATTTGTCCTTAATTTAGGTAGCT 1903  
Qy 263 GAAGGTGAAGATGATTTTATTGAAGAGGTATTTGTTAATGAAGATGATTCCTCAGTG 322  
Db 1902 GAAGGTGAAGATGATTTTATTGAAGAGGTATTTGTTAATGAAGATGATTCCTCAGTG 1843  
Qy 323 GAGGCTGTGACACACACCTAATACTCCCGAAACCTTCTGCTGATGAAATTTAGCAATCCA 382  
Db 1842 GAGGCTGTGACACACACCTAATACTCCCGAAACCTTCTGCTGATGAAATTTAGCAATCCA 1783  
Qy 383 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCT 442  
Db 1782 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCT 1723  
Qy 443 GTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCTGGAACAT 502  
Db 1722 GTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCTGGAACAT 1663  
Qy 503 TGGTCTGTATAGAGATATCTTGAATCTATGTACTTGAATCAAACTAACAGAAATTT 562  
Db 1662 TGGTCTGTATAGAGATATCTTGAATCTATGTACTTGAATCAAACTAACAGAAATTT 1603  
Qy 563 CATGTGCTATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCGCCAAAATTTAT 622  
Db 1602 CATGTGCTATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCGCCAAAATTTAT 1543  
Qy 623 GAATTTCTTAAAGTCAAAGAGGAGGAGTTCCAAGAAATATCTACAGAAATCTTTCGAGCAT 682  
Db 1542 GAATTTCTTAAAGTCAAAGAGGAGGAGTTCCAAGAAATATCTACAGAAATCTTTCGAGCAT 1483  
Qy 683 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGGTGGGAAACA 742  
Db 1482 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGGTGGGAAACA 1423  
Qy 743 CAATTTCTTGAATAGATCTACAGATCTAAATCTTGGGAAATTTATAAATCTGTTCTCT 802  
Db 1422 CAATTTCTTGAATAGATCTACAGATCTAAATCTTGGGAAATTTATAAATCTGTTCTCT 1363  
Qy 803 GGAATACTAATGAAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAAATTTCAATTAAT 862  
Db 1362 GGAATACTAATGAAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAAATTTCAATTAAT 1303  
Qy 863 TCTTGTGAGTCTCCAAAGCCTTAACCAAGTAGAGCCAGAACTGACCAATTTCTCAGCCCTACT 922  
Db 1302 TCTTGTGAGTCTCCAAAGCCTTAACCAAGTAGAGCCAGAACTGACCAATTTCTCAGCCCTACT 1243  
Qy 923 TCAGAAAACACACAGAAAGCTTTTCAATGATCTGTTTAAATAATAGCAAAACCGTCTGAA 982  
Db 1242 TCAGAAAACACACAGAAAGCTTTTCAATGATCTGTTTAAATAATAGCAAAACCGTCTGAA 1183  
Qy 983 AATACAGAGAGAAACCAAAATCAGAAATTTTATGAGAGTGAATGACTGTAGAGAGGTC 1042  
Db 1182 AATACAGAGAGAAACCAAAATCAGAAATTTTATGAGAGTGAATGACTGTAGAGAGGTC 1123

1043 TATGATTACCTGATGTATGTAGGAGGGGTAGTTTTCCAGGTTCTCTGACTGCTTCATCAT 1102  
1122 TATGATTACCTGATGTATGTAGGAGGGGTAGTTTTCCAGGTTCTCTGACTGCTTCATCAT 1063  
1103 CTCTTAATGGGAACCTGGAATCTCTTTAAACACCCCTGGAAATGATPACTGATTAAT 1162  
1062 CTCTTAATGGGAACCTGGAATCTCTTTAAACACCCCTGGAAATGATPACTGATTAAT 1003  
1163 CTTGAGTGTAAACTAGACAGCTATTTCCAGGAGCACCGTTTGGTCTCACTCATACACTT 1222  
1002 CTTGAGTGTAAACTAGACAGCTATTTCCAGGAGCACCGTTTGGTCTCACTCATACACTT 943  
1223 CTTGAGGATGCTATATCTCTGTAACCACTGAACTCGCTCTCTCCAGATAAGCAAAA 1282  
942 CTTGAGGATGCTATATCTCTGTAACCACTGAACTCGCTCTCTCCAGATAAGCAAAA 883  
1283 GGAGCAAAACAGACTTTTGAAGAAATGATTAATTCAGATCTCTTGTAGTCAAGTGT 1342  
882 GGAGCAAAACAGACTTTTGAAGAAATGATTAATTCAGATCTCTTGTAGTCAAGTGT 823  
1343 ATTGTTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGTATGGCTTACAGCAACCA 1402  
822 ATTGTTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGTATGGCTTACAGCAACCA 763  
1403 GTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATACAGGAACTGTTTCCA 1462  
762 GTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATACAGGAACTGTTTCCA 703  
1463 GAGCTCAATTAAGGTCAAAAGGAAGTATACCTCTGTGACATCTTGATGTAACACTTTGGA 1522  
702 GAGCTCAATTAAGGTCAAAAGGAAGTATACCTCTGTGACATCTTGATGTAACACTTTGGA 643  
1523 TTTGGTATAGAAATACCCATTTGAAATTTCTGCTGTGCGAGGGTGTGAAATTTACTTTT 1582  
642 TTTGGTATAGAAATACCCATTTGAAATTTCTGCTGTGCGAGGGTGTGAAATTTACTTTT 583  
1583 TTTGGTATATCTTATATATATATGATACATCGCTCTGAAATTTTATGTTATTTTGT 1642  
582 TTTGGTATATCTTATATATATATGATACATCGCTCTGAAATTTTATGTTATTTTGT 523  
1643 TTTTATAAAGACTAACCAAACTTAATGATTAATAAGTGAATGAGTCTCTAGTCTTTTCA 1702  
522 TTTTATAAAGACTAACCAAACTTAATGATTAATAAGTGAATGAGTCTCTAGTCTTTTCA 463  
1703 TTTGCTAGCTGTGATCCAAATTTTATAGAACATAAGTCACTTTGTTATGTCATTTTAA 1762  
462 TTTGCTAGCTGTGATCCAAATTTTATAGAACATAAGTCACTTTGTTATGTCATTTTAA 403  
1763 AAGAGAAATTCATAATGATGTTATGCGCAACAGATAGACTGATAAACTTCGTATTGTA 1822  
402 AAGAGAAATTCATAATGATGTTATGCGCAACAGATAGACTGATAAACTTCGTATTGTA 343  
1823 TAGCTTTGAAATAATTAATGCTAGTATGAGAGAAACAGAAATAGATCTGATTTCTTAG 1882  
342 TAGCTTTGAAATAATTAATGCTAGTATGAGAGAAACAGAAATAGATCTGATTTCTTAG 283  
1883 AGTTAATATATTTTAGTATGTTGTTTCTTTTATTTTATTTTATTTGACATAGTAACTGTG 1942  
282 AGTTAATATATTTTAGTATGTTGTTTCTTTTATTTTATTTTATTTGACATAGTAACTGTG 223  
1943 TATCTATAATAAGCATCTATATGAGTTTATGAGTTTATATAATAAAAA 1987  
222 TATCTATAATAAGCATCTATATGAGTTTATGAGTTTATATAATAAAAA 178

RESULT 3  
AA160816/c  
ID AA160816 standard; cDNA; 2176 BP.  
XX  
AC AA160816;  
XX  
DT 22-OCT-2001 (first entry)  
XX

DE Human polynucleotide SEQ ID NO 4805.  
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; Huntington's disease; haemostatic;  
KW Alzheimer's; Parkinson's disease; Shy-Drager Syndrome; chemotactic;  
KW amyotrophic lateral sclerosis; drug screening; arthritis; inflammation;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 13-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSR-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR P-PSDB; AAM41660.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Claim 1; SEQ ID NO 4805; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;

Query Match 92.4%; Score 1840.2; DB 22; Length 2176;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1842; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 143 TTCAGGACACACAGAAAGGGGAGAGATCATTTGGATCAGCAGATAGGTAGCAAAAT 202  
Db 2022 TTTCCGAAACACAGAAAGGGGAGAGATCATTTGGATCAGCAGATAGGTAGCAAAAT 1963  
Qy 203 AAAGGAGTATCAAAAGTACCAATGGAGGAGCTATGTCCTTAATTTATGGTGTAGCT 262  
Db 1962 AAAGGAGTATCAAAAGTACCAATGGAGGAGCTATGTCCTTAATTTATGGTGTAGCT 1903  
Qy 263 GAAGGTGAAGATGATTTTATTTGAGAGGATTTGTTGTTGATGAGATGATTTCCAGTG 322

1902 GAAGGTGAAGATGATTTTATTGAAGAAGGTATTTGTTGAATGGAAGATGATTTCTCCAGTG 1843  
323 GAGGCTGTGAGCACACCTAACTACTCCCGAAGACCTTGTGTCATGGAAATTTAGCATTTCCA 382  
1842 GAGGCTGTGAGCACACCTAACTACTCCCGAAGACCTTGTGTCATGGAAATTTAGCATTTCCA 1783  
383 TATGTAGACTTTTGTGAGGATCCCTCCTCTGAAAGGAAGGAGAAAGAAAGAAATTCCT 442  
1782 TATGTAGACTTTTGTGAGGATCCCTCCTCTGAAAGGAAGGAGAAAGAAAGAAATTCCT 1723  
443 GTGTTTGTATGATGTTGAAGAAATGATAGAGAGAGAGTGGACAGAGCCCTGAACAT 502  
1722 GTGTTTGTATGATGTTGAAGAAATGATAGAGAGAGAGTGGACAGAGCCCTGAACAT 1663  
503 TGGTCTGTCTATAGAGATATCTTGAATCTATGTTGATCAAACTAACAGAAATTT 562  
1662 TGGTCTGTCTATAGAGATATCTTGAATCTATGTTGATCAAACTAACAGAAATTT 1603  
563 CATGGTGCAATTCCTGATGCCAGCTTCTTCAAGAGAGATCATTTGGCCGCCAAATTTAT 622  
1602 CATGGTGCAATTCCTGATGCCAGCTTCTTCAAGAGAGATCATTTGGCCGCCAAATTTAT 1543  
623 GAATTCCTTAAGTCAAGAGAGGAGAGTTCAGAGATATCTACAGAACTTCTGCGAGCAT 682  
1542 GAATTCCTTAAGTCAAGAGAGGAGAGTTCAGAGATATCTACAGAACTTCTGCGAGCAT 1483  
683 CCGAAGTCAAGTAAATGATCTTCTGCGAGATCTTCTTCCCTAAATGGTGGGGAACA 742  
1482 CCGAAGTCAAGTAAATGATCTTCTGCGAGATCTTCTTCCCTAAATGGTGGGGAACA 1423  
743 CAATTCCTTGAATGATATCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCT 802  
1422 CAATTCCTTGAATGATATCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCT 1363  
803 GGAAACTAATGAAGAGAGAGTTCAGCAATTTGGAACTTTTATCATGAATTTCAATTAAT 862  
1362 GGAAACTAATGAAGAGAGAGTTCAGCAATTTGGAACTTTTATCATGAATTTCAATTAAT 1303  
863 TCTTGTGATCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACT 922  
1302 TCTTGTGATCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACT 1243  
923 TCAGAAACCAAGAGAGAGTTCATGATCTGTTTAAATAATGCAACCGCTGAA 982  
1242 TCAGAAACCAAGAGAGAGTTCATGATCTGTTTAAATAATGCAACCGCTGAA 1183  
983 AATACAGAGAGAGAGAGAGTTCATGATCTGTTTAAATAATGCAACCGCTGAA 1042  
1182 AATACAGAGAGAGAGAGAGTTCATGATCTGTTTAAATAATGCAACCGCTGAA 1123  
1043 TATGATTAATCTGATGATGAGAGAGGAGTTCAGGATTCCTGAGCTTCCATCAT 1102  
1122 TATGATTAATCTGATGATGAGAGAGGAGTTCAGGATTCCTGAGCTTCCATCAT 1063  
1103 CTCTTAATGGAACTCGAATCTCTTTTAAACCAACCTGGAATGATGATGATGAT 1162  
1062 CTCTTAATGGAACTCGAATCTCTTTTAAACCAACCTGGAATGATGATGATGAT 1003  
1163 CTTCAGTGTAACTAGACAGAGTATTTTCAGGAGAGAGTTCCTGCTCACTATACAT 1222  
1002 CTTCAGTGTAACTAGACAGAGTATTTTCAGGAGAGAGTTCCTGCTCACTATACAT 943  
1223 CTACAGAGATGCTATATCTTGTGAAACACTGAACTGCTCTCTCCAGATTAAGCAAAA 1282  
942 CTACAGAGATGCTATATCTTGTGAAACACTGAACTGCTCTCTCCAGATTAAGCAAAA 883  
1283 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATCCAGATCTGTTAGTCAAGTGT 1342  
882 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATCCAGATCTGTTAGTCAAGTGT 823  
1343 ATTGGTGAAGAAACCAAGATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1402  
822 ATTGGTGAAGAAACCAAGATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 763

1403 GTACTCAACAGCAGCTGACTTATGTTTATTTGGCAATTTGTGATACAGGAACCTGTTTCCA 1462  
762 GTACTCAACAGCAGCTGACTTATGTTTATTTGGCAATTTGTGATACAGGAACCTGTTTCCA 703  
1463 GAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAACACTTGA 1522  
702 GAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAACACTTGA 643  
1523 TTTGGTATAGATTAACCCATTTGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 1582  
642 TTTGGTATAGATTAACCCATTTGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 583  
1583 TTTGGTATATCTTAT 1642  
582 TTTGGTATATCTTAT 523  
1643 TTTTAATAAGACTAACACAACTTAATGATTAATAAGTGAAGTCTCATAGTCTTTTCA 1702  
522 TTTTAATAAGACTAACACAACTTAATGATTAATAAGTGAAGTCTCATAGTCTTTTCA 463  
1703 TTTGCTAGCTGTGATCAAAATTTTATAGACATTAAGTCACTTGTATTGTCATTTTAA 1762  
462 TTTGCTAGCTGTGATCAAAATTTTATAGACATTAAGTCACTTGTATTGTCATTTTAA 403  
1763 AAGAGAAATTTCAATATGATTTATGCGCAACAGATAAGACTGATAAACTTCGTATTGTA 1822  
402 AAGAGAAATTTCAATATGATTTATGCGCAACAGATAAGACTGATAAACTTCGTATTGTA 343  
1823 TAGCTTTGAAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1882  
342 TAGCTTTGAAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 283  
1883 AGTTAATAATTTTAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1942  
282 AGTTAATAATTTTAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 223  
1943 TATCTATAATAAGCATCTCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 1987  
222 TATCTATAATAAGCATCTCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 178

## RESULT 4

AAI59029

ID AAI59029 standard; cdna; 1716 BP.

XX AC AAI59029;

XX AC AAI59029;

DT 22-OCT-2001 (first entry)

XX XX

Human polynucleotide SEQ ID NO 1232.

DE DE

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Human; nontopic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-052317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

PR	19-OCT-2000; 2000US-0693036.
XX	29-NOV-2000; 2000US-0727344.
XX	(HYSE-) HYSEQ INC.
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Pi Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	WPI; 2001-442253/47.
DR	P-PSDB; AAM39873.
XX	Novel nucleic-acids and polypeptides, useful for treating disorders
XX	such as central nervous system injuries -
PT	Claim 1; SEQ ID NO 1232; 10078pp; English.
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM39842-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;
SQ	Query Match 76.9%; Score 1532.8; DB 22; Length 1716;
	Best Local Similarity 99.9%; Pred. No. 5.2e-304;
	Matches 1534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	143 TTCAAGGACACACAGAAAAGGGGGAATCATTTTGGNAATCAGCAGATAAGGTAGCAAAATT 202
DB	173 TTTCGGAACACACAGAAAAGGGGGAATCATTTTGGNAATCAGCAGATAAGGTAGCAAAATT 232
QY	203 AAAGGAGTAGTAATCAAAGTAGTACCAGTAATGAGGAGCGATGTTGCCCTAAATATGGTGAGCT 262
DB	233 AAAGGAGTAGTAATCAAAGTAGTACCAGTAATGAGGAGCGATGTTGCCCTAAATATGGTGAGCT 292
QY	263 GAAGGTGAAGATGATTTTATGAAGAAGTATGTTGTAATGGAAGATGATTCCTCAGTG 322
DB	293 GAAGGTGAAGATGATTTTATGAAGAAGTATGTTGTAATGGAAGATGATTCCTCAGTG 352
QY	323 GAGGCTGTGAGCACCTTAATATCTCCCGAACCTTGCCTGCAATGGAAATTAGCATTTCCA 382
DB	353 GAGGCTGTGAGCACCTTAATATCTCCCGAACCTTGCCTGCAATGGAAATTAGCATTTCCA 412
QY	383 TATGTAGACTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGGAAAAAGAAAGAAATTCCT 442
DB	413 TATGTAGACTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGGAAAAAGAAAGAAATTCCT 472
QY	443 GTGTTTTGTATGATGTTGAAGAAATGATPAGAAGAGCAGTTGGACACGAGCCTGAACAT 502
DB	473 GTGTTTTGTATGATGTTGAAGAAATGATPAGAAGAGCAGTTGGACACGAGCCTGAACAT 532
QY	503 TGCTCTGTCTATGAGATATCTTGAATCTATGTAATCTGATCAAACTAACAGAAATTT 562
DB	533 TGCTCTGTCTATGAGATATCTTGAATCTATGTAATCTGATCAAACTAACAGAAATTT 592
QY	563 CATGGTGATTTCTCTGATGCGCAGCTTCCTTCTPAAGAGGATCATTTGGCCCCCAAAATTTAT 622

## RESULT 5

623 GAATTCTTAAAGTCAAAGAGGGAGAGTTCCAAGAAATATCTACAGAAACTTCTGCAGCAT 682

Qy



	ABO79518	
ID	ABQ7951B standard; cDNA; 3145 BP.	
XX		
AC	ABQ7951B;	
XX		
DT	25-NOV-2002 (first entry)	
XX		
DE	cDNA encoding a protein similar to human sorting nexin.	
XX		
KW	Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;	
XW	SSH 6; gene; human; nExin; ss.	
XX		
OS	Homo sapiens.	
XX		
TH	Key Location/Qualifiers	
FT	CDS 128..2968	
FT	/tag= a	
XX		
XX	WO200262839-A2.	
PX		
FN	15-AUG-2002.	
PD		
XX		
PF	05-FEB-2002; 2002WO-BP01327.	
XX		
XX	07-FEB-2001; 2001EP-0200439.	
PR	(UTMA-) UNIV MAASTRICHT.	
PA		
XX	Daemen MJAP, Cleutjens CBJM, Zaman GJR;	
XX	WPI; 2002-643400/69.	
PI	P-PSDB; ABB81193.	
DR		
DR	Use of a polynucleotide differentially expressed in ruptured and stable	
XX	atherosclerotic plaques as a marker for atherosclerosis, useful in	
PT	treating, diagnosing or preventing atherosclerosis -	
XX		
PS	Example 6; Page 35-37; 44pp; English.	
XX		
CC	The invention relates to the use of a polynucleotide differentially	
CC	expressed in ruptured and stable atherosclerotic plaques as a marker	
CC	for atherosclerosis, where the polynucleotides can be selected from the	
CC	sequences shown in ABQ79517-19. The polynucleotides are useful as a	
CC	marker of atherosclerosis, which may be used: (i) in the diagnosis,	
CC	prevention and treatment of atherosclerosis; (ii) as serum/plasma markers	
CC	to screen patients at risk for plaque instability to evaluate the effects	
CC	of other treatments; (iii) in the preparation of vector molecules for the	
CC	expression of the encoded protein in host cells; and (iv) in the	
CC	identification of functional targets or analogues of the gene. The	
CC	polynucleotides, the encoded proteins or antibodies against the proteins	
CC	may be used to target other therapeutics to an unstable plaque.	
CC	Modulation of the expression of the polynucleotide can increase plaque	
CC	stability and therefore inhibit the progression of atherosclerotic	
CC	cardiovascular disease. Modulators may be used to prepare pharmaceuticals	
CC	for atherosclerotic disorders. The present sequence represents a cDNA	
CC	encoding a protein similar to human sorting nexin.	
XX		
SQ	Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;	
	Query Match 75.5%; Score 1504; DB 24; Length 3145;	
	Best Local Similarity 99.5%; Pred. No. 4.6e-298;	
	Matches 1529; Conservative 0; Mismatches 5; Indels 2; Gaps 2;	
Qy	143 TTCAAGAACACACAGAAAAGGGGAGATCATTTTGAATCAGCAGATAGGTAGCAAAATT 202	
Dd	1598 TTTTCGGAACACACAGAAAAGGGGAGATCATTTTGAATCAGCAGATAGGTAGCAAAATT 1657	
Qy	203 AAAGGAGTATCAAAGTACCAATCGGGAGCTATGTCCTAATTAATGCTAGCT 262	
Dd	1658 AAAGGAGTATTCAGAAGTACCAATCGGGAGCTATGTCCTAATTAATGCTAGCT 1717	
Qy	263 GAAGGTGAAGATGATTTTATTAAGAGAGGATTTGTTGAATGAAGATGATTTCTCCAGTG 322	

1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGTGATACAGGAACTGTTTCCA 1462  
 2858 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGTGATACAGGAACTGTTTCCA 2917  
 1463 GAGCTCAATAAGGTACAAAGGAAAGTTACCTCTCTGTGACATCTTGGATGTAACACACTTGGGA 1522  
 2918 GAGCTCAATAAGGTACAAAGGAAAGTTACCTCTCTGTGACATCTTGGATGTAACACACTTGGGA 2977  
 1523 TTGTGTATAGTAATACCCATTGAAATTTCTCTCTGTGCGAGGGTGGTGAATAATTTACTTTT 1582  
 2978 TTGTGTATAGTAATACCCATTGAAATTTCTCTCTGTGCGAGGGTGGTGAATAATTTACTTTT 3037  
 1583 TTGGGTATATCTTAT 1642  
 3038 TTGGGTATATCTTAT 3096  
 1643 TTTTAAATTAAGCACTAACCAACCACTTAATGATTAAAA 1678  
 3037 TTTTAAATTAAGCACTAACCACTTAATGATTAAAA 3131

RESULT 6

AAI59030  
 ID AAI59030 standard; cDNA; 1551 BP.  
 AC AAI59030;  
 DT 22-OCT-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 1233.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amocrophic/lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.

OS Homo sapiens.  
 XX WO200153312-A1.  
 XX 26-JUL-2001.  
 XX 26-DEC-2000; 2000WO-US34263.  
 XX 21-JAN-2000; 2000US-0488725.  
 XX 25-APR-2000; 2000US-0552317.  
 XX 09-JUL-2000; 2000US-0598042.  
 XX 19-JUL-2000; 2000US-0620312.  
 XX 03-AUG-2000; 2000US-0653450.  
 XX 14-SEP-2000; 2000US-0682191.  
 XX 19-OCT-2000; 2000US-0693036.  
 XX 29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX PI Zhao J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX WPI; 2001-442253/47.  
 XX P-FSD; AM39874.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 XX such as central nervous system injuries -

PS Claim 1; SEQ ID NO 1233; 10078pp; English.  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AA42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;

Query Match 59.9%; Score 1192.8; DB 22; Length 1551;  
 Best Local Similarity 89.1%; Pred. No. 1.7e-234;  
 Matches 1369; Conservative 0; Mismatches 2; Indels 165; Gaps 1;

QY	143	TTTCAGGACACACAGAAAGGGGAGCAATCAATTTGGAATCAGCAGAAATAGGTAGCAAAATT	202
DB	173	TTTCGGAACACACAGAAAGGGGAGCAATCAATTTGGAATCAGCAGAAATAGGTAGCAAAATT	232
QY	203	AAAGGAGTATTCAAAAGTACCAATGAGGAGGAGTATGTTGCCCTAAATATGTTGTTAGCT	262
DB	233	AAAGGAGTATTCAAAAGTACCAATGAGGAGGAGTATGTTGCCCTAAATATGTTGTTAGCT	292
QY	263	GAAGGTGAAGATGATTTTATTGGAAGAGTATGTTGTAATGGAAGATGATTTCTCCAGTG	322
DB	293	GAAGGTGAAGATGATTTTATTGGAAGAGTATGTTGTAATGGAAGATGATTTCTCCAGTG	352
QY	323	GAGGCTGTGAGACACACCTTAATCTCCGAAACCTTGTGATGGAATAGCAATTTTCCCA	382
DB	353	GAGGCTGTGAGACACCTTAATCTCCGAAACCTTGTGATGGAATAGCAATTTTCCCA	412
QY	383	TATGTAGACTTTTGTGAGATCCCTCTCTGAAAGAGGAGAGAGAGAGAGAGAGATTTCT	442
DB	413	TATGTAGACTTTTGTGAGATCCCTCTCTGAAAGAGGAGAGAGAGAGAGAGATTTCT	472
QY	443	GTGTTTGTATGATGTTGAAAGAAATGATGAAGAGAGAGTGTGACAGAGCTGAACAT	502
DB	473	GTGTTTGTATGATGTTGAAAGAAATGATGAAGAGAGAGTGTGACAGAGCTGAACAT	532
QY	503	TGCTCTGTCTATGAAGATATCTTGAATTTCTATGTTGATCAAACTAACAGATTT	562
DB	533	TGCTCTGTCTATGAAGATATCTTGAATTTCTATGTTGATCAAACTAACAGATTT	592
QY	563	CATGTTGCATTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTAAT	622
DB	593	CATGTTGCATTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTAAT	652
QY	623	GAATTTCTTAAAGTCAAGAGGAGAGAGTTCGAAGATATCTACAGAACTTTCTGCAGCAT	682
DB	653	GAATTTCTTAAAGTCAAGAGGAGAGAGTTCGAAGATATCTACAGAACTTTCTGCAGCAT	712
QY	683	CCAGAACTGAGTAAATAGTCAACTTCTGGCAGACTTTTCTTCCCTTAATGGTGGGAAACA	742
DB	713	CCAGAACTGAGTAAATAGTCAACTTCTGGCAGACTTTTCTTCCCTTAATGGTGGGAAACA	772
QY	743	CAATTTCTGATGAAGATATCTACAGATGTAATCTTTGGGAATTAATAAATCTGTTCT	802
DB	773	CAATTTCTGATGAAGATATCTACAGATGTAATCTTTGGGAATTAATAAATCTGTTCT	832
QY	803	GGAAACTTAATGAAGAGAGAGGTCAGCATTTTGGAACTTTTATCATGAATTTTCAATTAAT	862
DB	833	GGAAACTTAATGAAGAGAGAGGTCAGCATTTTGGAACTTTTATCATGAATTTTCAATTAAT	892
QY	863	TCTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTACCACTTCTCAGCCCTACT	922
DB	893	TCTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTACCACTTCTCAGCCCTACT	952
QY	923	TCAGAAACCAACAGAGAGCTTTTCAATGATCTCTTTTAAATAATTAATGCAAAACCGTGTGAA	982

953 TCAGAAACCAACAGAGCTTTCAATGATCTGTTTAAATAATATGCAACCGTGTGAA 1012  
983 AATACAGAGAGAGCAACCAATCAGATTTATTTATGGAGGTGATGACTGTAGAGAGGATC 1042  
1013 AATACAGAGAGAGCAACCAATCAGATTTATTTATGGAGGTGATGACTGTAGAGAGGATC 1072  
1043 TATGATTACCTGATGATGTAGGAGCGGTAGTTTTCCAGGTTCTCTGACTGGCTTCATCAT 1102  
1073 TATGATTACCTGATGATGT----- 1092  
1103 CTCCTTAATGGGAACCTCGAATCCTCTTTAAACACCCCTGGAAATATGATGATTACTAT 1162  
1093 ----- 1092  
1163 CTTCTAGTGTAACTAGACAGCTATTTTCAGGAGCACCGTTTGGTCTCTACTATACACTT 1222  
1093 ----- 1092  
1223 CTCAGAGATGCTATATCTGTGAAACACTGAACTCGCTCTCTCCAAAGATAAGCAAAA 1282  
1093 -----AGATGCTATATCTGTGAAACACTGAACTCGCTCTCTCCAAAGATAAGCAAAA 1147  
1283 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTCTGTAGTCAAGTGT 1342  
1148 GGGGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTCTGTAGTCAAGTGT 1207  
1343 ATTGGTGAAGAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1402  
1208 ATTGGTGAAGAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1267  
1403 GTACTACACAGAGCTGACTATGTTTTTATGGACATTTGATACAGCAAGTCTTTTCCA 1462  
1268 GTACTACACAGAGCTGACTATGTTTTTATGGACATTTGATACAGCAAGTCTTTTCCA 1327  
1463 GAGCTCAATAGGTACAAAGGAAGTTACTCTGTGACATCTGGAATGAAACACTTGA 1522  
1328 GAGCTCAATAGGTACAAAGGAAGTTACTCTGTGACATCTGGAATGAAACACTTGA 1387  
1523 TTGGGTATAGTAACCACTTGAATTTCTGCTGTCGGGGTGGTGAATTTACTTTT 1582  
1388 TTGGGTATAGTAACCACTTGAATTTCTGCTGTCGGGGTGGTGAATTTACTTTT 1447  
1583 TTGGGTATATCTTAT 1642  
1448 TTGGGTATATCTTAT 1507  
1643 TTTTATATAGACTACACAACTTAATGATTAATAA 1678  
1508 TTTTATATAGACTACACAACTTAATGATTAATAA 1543

RESULT 7  
AAZ17379  
ID AAZ17379 standard; cDNA; 779 BP.  
XX  
AC AAZ17379;  
XX  
DT 12-OCT-1999 (first entry)  
XX  
DE Human gene expression product cDNA sequence SEQ ID NO:4852.  
XX  
KW Human; gene; gene expression product; diagnosis; therapy; probe;  
KW detection; mapping; tissue typing; profiling; forensic; cancer;  
KW Genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09938972-A2.  
XX  
PD 05-AUG-1999.  
XX  
PF 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.  
PR 28-JAN-1998; 98US-0072910.  
PR 24-FEB-1998; 98US-0075954.  
PR 31-MAR-1998; 98US-0080114.  
PR 03-APR-1998; 98US-0080515.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I,  
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
XX WPI; 1999-494092/41.  
DR  
XX Novel human genes and their expression products which are  
PT differentially expressed in different cell types  
PT  
XX Claim 1; Page 2307; 2479pp; English.  
XX  
CC The present invention describes a library of human polynucleotides  
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
CC a method of detecting differentially expressed genes correlated with the  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample from a cell  
CC suspected of being cancerous, where the gene product is encoded by one  
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
CC polynucleotides can be used as a source of primers and probes, which can  
CC be used for a variety of purpose, e.g. detection of expression levels,  
CC mapping, tissue typing or profiling, forensics, genetic analysis and  
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
CC can be used for raising antibodies for experimental, diagnostic and  
CC therapeutic purposes. The polynucleotides may also be used to construct  
CC arrays for diagnostics (which may be used to determine function of an  
CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.  
XX  
SQ Sequence 779 BP; 248 A; 143 C; 153 G; 232 T; 3 other;  
Query Match 32.9%; Score 655.4; DB 20; Length 779;  
Best Local Similarity 98.3%; Pred. No. 1.2e-124;  
Matches 704; Conservative 0; Mismatches 7; Indels 5; Gaps 4;  
QY 882 CTAAACCAAGTAGACACAG-AACTGACCATTTCTGAGCCCTACTTCAGAAACCAACAGAG 940  
Db 59 CGAGACCAAGTAGACACAGAACTGACCATTTCTCAGTCTCTACTTCAGAAACCAACAGAG 118  
QY 941 CTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATATACAGAGAGAAAGCAA 1000  
Db 119 CTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATATACAGAGAGAAAGCAA 178  
QY 1001 AATCAGAAATATTTTATGGAGGTGATGACTGTAGAGAGGTCTATGATTACCTGATGAT 1060  
Db 179 AATCAGAAATATTTTATGGAGGTGATGACTGTAGAGAGGTCTATGATTACCTGATGAT 238  
QY 1061 GTAGACCGGTAGTTTTCAGGTTCCGACTGTGCTTATCTCTTAAATGGGAACTCGA 1120  
Db 239 GTAGACCGGTAGTTTTCAGGTTCCGACTGTGCTTATCTCTTAAATGGGAACTCGA 298  
QY 1121 ATCTCTTTTAAACACCCCTGGAATGTATGATTACTTCTTCAGTGTAAACTAGAA 1180  
Db 299 ATCTCTTTTAAACACCCCTGGAATGTATGATTACTTCTTCAGTGTAAACTAGAA 358  
QY 1181 CAGCTATTTCAGGAGCACCGTTTGGTCTCACTCAACACTCTCAGAGATGCTATATTC 1240

Db 359 CAGCTATTTCAGGAGCACCCTGGTGGTCTCACTCATACACTTCTCAGAGATGCTATATTC 418  
Qy 1241 TGTGAACAACCTGAACCTGCTCTCTCCAAAGATAAGCAAAAGAGCAAAACAGACTTTT 1300  
Db 419 TGTGAACAACCTGAACCTGCTCTCTCCAAAGATAAGCAAAAGAGCAAAACAGACTTTT 478  
Qy 1301 GAAGAAATGATGATTAATACATCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAG 1360  
Db 479 GAAGAAATGATGATTAATACATCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAG 538  
Qy 1361 TATGAAGAGCATCAGACTTCTGTTGATGGCTTACAGCAACAGTACTTCAACAGAGAGCTG 1420  
Db 539 TATGAAGAGCATCAGACTTCTGTTGATGGCTTACAGCAACAGTACTTCAACAGAGAGCTG 598  
Qy 1421 ACTTATGTTTATTTGGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1480  
Db 599 ACTTATGTTTATTTGGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 658  
Qy 1481 AAGGAAGTACTCTGTGACATCTTTGGGATGATGATGATGATGATGATGATGATGATGATGAT 1537  
Db 659 AAGGAAGTACTCTGTGACATCTTTGGGATGATGATGATGATGATGATGATGATGATGATGAT 718  
Qy 1538 CCCATTGAAATTTCTGCTGCTG-CCAGGAGTGTAGAAATTTACTTTTTTGGGTATAT 1592  
Db 719 CCCATTGAAATTTCTGCTGCTG-CCAGGAGTGTAGAAATTTACTTTTTTGGGTATAT 774

RESULT 8  
ABL63556/c  
ID ABL63556 standard; DNA; 451 BP.  
AC  
AC ABL63556;  
XX  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Breast cancer related gene sequence SEQ ID NO:1893.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.

28-SEP-2000; 2000US-236033P.  
28-SEP-2000; 2000US-236034P.  
28-SEP-2000; 2000US-236109P.  
28-SEP-2000; 2000US-236111P.  
29-SEP-2000; 2000US-236842P.  
29-SEP-2000; 2000US-236891P.  
02-OCT-2000; 2000US-237172P.  
02-OCT-2000; 2000US-237173P.  
02-OCT-2000; 2000US-237279P.  
02-OCT-2000; 2000US-237294P.  
02-OCT-2000; 2000US-237295P.  
02-OCT-2000; 2000US-237316P.  
03-OCT-2000; 2000US-237425P.  
03-OCT-2000; 2000US-237598P.  
03-OCT-2000; 2000US-237604P.  
03-OCT-2000; 2000US-237606P.  
03-OCT-2000; 2000US-237608P.  
01-NOV-2000; 2000US-244867P.  
01-NOV-2000; 2000US-245084P.  
(AVAL-) AVALON PHARM.  
PA  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-189264/24.  
DR  
PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
XX Claim 1; SEQ ID 1893; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
XX anti-neoplastic agent. The method involves exposing cells to a chemical  
XX agent to be tested for anti-neoplastic activity, determining a change in  
XX expression of at least one gene (I) of a signature gene set, where (I)  
XX comprises a sequence (S) selected from 847 sequences (given in ABL61664  
XX to ABL70110), or is at least 95% identical to (S), where a change in  
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic  
XX activity and can be used in gene therapy. M1 can be used for screening  
XX an anti-neoplastic agent, and can be used for producing a product which  
XX is the data collected with respect to the anti-neoplastic agent as a  
XX result of M1, and the data is sufficient to convey the chemical  
XX structure and/or properties of the agent. M1 can be used in the  
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,  
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
XX carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 451 BP; 179 A; 73 C; 52 G; 146 T; 1 other;  
Query Match 21.4%; Score 427.2; DB 24; Length 451;  
Best Local Similarity 98.9%; Pred. No. 4.8e-78;  
Matches 440; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
Qy 1538 CCCATTGAAATTTCTGCTGCTGCGA-GGGTGGTAGAAATTTACTTTTTTGGGTATATTTCT 1596  
Db 445 CCCATGGAATTCGGCTGTGGAGGGGGTGGTAGAAATTTACTTTTTTGGGTATATTTCTN 386  
Qy 1597 ATATATATTTATGTACATCGCTCTCGAAATTTAGTATTTTCTTTTAAAGACT 1656  
Db 385 ATATATATTTATGTACATCGCTCTCGAAATTTAGTATTTTCTTTTAAAGACT 326  
Qy 1657 AACACAACTTAAATGATTAAGAGTGAAGTCTCATAGTCTTTTCTTTAGCTGTGA 1716  
Db 325 AACACAACTTAAATGATTAAGAGTGAAGTCTCATAGTCTTTTCTTTAGCTGTGA 266  
Qy 1717 TCCAAATTTTATAGACATAAGTCACTGTTTATTTGCAATTTTAAAGAGAAATTCAT 1776  
Db 265 TCCAAATTTTATAGACATAAGTCACTGTTTATTTGCAATTTTAAAGAGAAATTCAT 206



PR 17-DEC-1998; 98US-113678P.  
PR 15-DEC-1999; 99US-0465231.  
XX  
PA (BYAT//) BYATT J C.  
PA (MATH//) MATHALAGAN N.  
PA (TAON//) TAO N.  
PA (WARR//) WARREN W C.  
XX  
PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX  
XX WPI; 2003-102386/09.  
XX  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
PT identification and analysis, cattle breeding or preparation of  
PT constructs for cattle gene expression and genetically improved cattle -  
XX  
PS Claim 2; SEQ ID No 540; 38pp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
XX lactation or muscle and fat deposition (designated LMFD), derived  
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 5912 nucleotide  
CC sequences, appearing as ABX50072-ABX5983, or complements of them.  
CC Also included are; (1) a transformed cell having a nucleic acid  
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMFD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 5912 bovine  
CC LMFD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX  
XX Sequence 402 BP; 141 A; 71 C; 76 G; 113 T; 1 other;  
XX  
XX Query Match 18.3%; Score 364.2; DB 25; Length 402;  
XX Best Local Similarity 94.0%; Pred. No. 3.6e-65;  
XX Matches 378; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
XX  
QY 726 CTAATGGTGGGAAACACAAATTTCTTGATAAGATACTACAGATGTAATCTTGGGAAA 785  
DB 1 CCATGGTGGGAAACACAGTTCTTGATAAGATACTACAGATGTAATCTTGGGAAA 60  
QY 786 TTATAAATCTGTTCTGGAAACATTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTA 845  
DB 61 TTATAAATCTGTTCTGGAAACATTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTCA 120  
QY 846 TCATGAATTTCAATTAATCTTGAGTCTCCAAAGCCTAAACCAAGTAGACCAAGCTGA 905  
DB 121 TCATGAATTTCAATTAATCTTGAGTCTCCAAAGCCTAAACCAAGTAGACCAAGCTGA 180  
QY 906 CCATTTCTGAGCCCTACTCTCAGAAACACACAGAGCTTTTCAATGATCTGTTTAAATA 965  
DB 181 CCATTTCTGAGCCCTACTCTCAGAGATAATAGAGAGCTTTTCAATGATCTGTTTAAATA 240  
QY 966 ATGCAAAACCGTCTCGAAATAACAGAGAGAGCAAAATCAGAAATTTTATGAGGTGA 1025  
DB 241 ATGCAAAACCGTCTCGAGATAACAGAGAGAGCAAAATCAGAAATTTTATGAGGTGA 300  
QY 1026 TGACTGTAGAGAGGTCTTANGANTTACCTGATGTATGTAGGACGGGTAGTTTCCAGGTT 1085

DB 301 TGACTGTAGAGAGGTCTTATGATTTACCTGATGTATGTATGACAGGTGTTTCCAGTTC 360  
QY 1086 CTGACTGCTTCATCATCTCTTAAATGGAACATCGAATCTCT 1127  
DB 361 CTGACTGCTTCATCATCTCTTAAATGGAACATCGAATCTCT 402  
RESULT 11  
AAC02755  
ID AAC02755 standard; cDNA; 358 BP.  
XX  
XX AAC02755;  
XX AC  
XX DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 2753.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX OS  
XX PN EP1033401-A2.  
XX PD 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX P-PSDB; AAG02749.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
XX Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;  
XX  
XX Query Match 17.9%; Score 357; DB 21; Length 358;  
XX Best Local Similarity 100.0%; Pred. No. 1e-63;  
XX Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 760 ACTACAGATGTAATCTTGGGAAATATATAAATCTCTTCTCGAAACATTAAGAGA 819  
DB 1 ACTACAGATGTAATCTTGGGAAATATATAAATCTCTTCTCGAAACATTAAGAGA 60  
QY 820 GAAAGTCAGCATTTGGAAACCTTTTATCATGAATTTCAATTAATCTTGTGAGTCTCCAAA 879  
DB 61 GAAAGTCAGCATTTGGAAACCTTTTATCATGAATTTCAATTAATCTTGTGAGTCTCCAAA 120

QY 880 GCTAAACCAAGTAGACCAAGCTGACCAATTCACGCCCTTCTCAGAAAACCAACAGAA 939  
 DB 121 GCTAAACCAAGTAGACCAAGCTGACCAATTCACGCCCTTCTCAGAAAACCAACAGAA 180  
 QY 940 GCTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAATACAGAGGAAGCA 999  
 DB 181 GCTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAATACAGAGGAAGCA 240  
 QY 1000 AATACAGAAATATTTATGAGGTGATGACTGAGAGGAGTCTATGATTACTGATGTA 1059  
 DB 241 AATACAGAAATATTTATGAGGTGATGACTGAGAGGAGTCTATGATTACTGATGTA 300  
 QY 1060 TGTAGGAGGGTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAAATGGGAAC 1116  
 DB 301 TGTAGGAGGGTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAAATGGGAAC 357

## RESULT 12

ABX50426  
 ID ABX50426 standard; cDNA; 305 BP.

AC ABX50426;

XX 25-FEB-2003 (first entry)

DT Bovine EST associated with lactation/muscle/fat deposition #355.

DE Bovine, ss; EST, expressed sequence tag; lactation; LMFD;  
 KW muscle deposition; fat deposition; genome mapping; gene identification;  
 KW gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137160-A1.

XX 26-SEP-2002.

XX 26-OCT-2001; 2001US-0983965.

XX 17-DEC-1998; 98US-113678P.

XX 15-DEC-1999; 99US-0465231.

XX (BYATT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

PI Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-102386/09.

DR Purified nucleic acid molecules, useful for genome mapping, gene  
 XX identification and analysis, cattle breeding or preparation of  
 XX constructs for cattle gene expression and genetically improved cattle -

PS Claim 2; SEQ ID No 355; 38pp; English.

XX The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMFD), derived  
 CC from cattle, and the LMFD nucleic acid can specifically hybridize to a  
 CC second nucleic acid molecule comprising any of 5912 nucleotide  
 CC sequences, appearing as ABX50072-ABX55983, or complements of them.  
 CC Also included are; (1) a transformed cell having a nucleic acid  
 CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the

CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMFD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 5912 bovine  
 CC LMFD EST (expressed sequence tag) nucleic acids.

CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?DocID=20020137160.

XX Sequence 305 BP; 109 A; 60 C; 51 G; 85 T; 0 other;

Query Match 13.4%; Score 267.2; DB 25; Length 305;

Best Local Similarity 92.4%; Pred. No. 2.3e-45;

Matches 281; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 653 CAAGTAATATCTACAGAACTTCTGCAGCATCCAGAACTAGTATAGTCAACTTCTGGCA 712

DB 1 CAGGAATATCTGCAGAACTTCTGCAACATCCAGATTGAGTAACAGTCAACTTCTGGCA 60

QY 713 GACTTTCTTCCCTTAATGTTGGGAAACACAATTTCTTGTATAGATACACAGATGTA 772

DB 61 GATTTCTTCTCCCAATGTTGGGAAACACAGTTTCTTGTATAGATACACAGATGTA 120

QY 773 AATCTTGGGAAATTAATAAATCTGTTCTCGGAAACTAATGAAGAGAAAGTTCAGCAT 832

DB 121 AATCTTGGGAAATTAATAAATCTGTTCTCGGAAACTAATGAAGAGAAAGTTCAGCAT 180

QY 833 TTGGAACCTTTTATCATGAATTCATTAATTTCTGTGAGTCTCCAAAGCCTTAAACCAAGT 892

DB 181 TTGGAACCTTTTATCATGAATTCATTAATTTCTGTGAGTCTCCAAAGCCTTAAACCAAGT 240

QY 893 AGACCAGAACTGACCATCTTCAGCCCTTCTCAGAAACCAACAGAAAGCTTTTCAATGAT 952

DB 241 AAACAGAACTGACCAATCTCAGCCCTTACATCAGAGTAATAACAGCTTTTATGAT 300

QY 953 CTGT 956

DB 301 CTGT 304

## RESULT 13

AAH69552  
 ID AAH69552 standard; cDNA; 311 BP.

XX AC AAH69552;

XX 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 826.

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

OS Homo sapiens.

XX WO200142467-A2.

PD 14-JUN-2001.

PF 08-DEC-2000; 2000WO-US33312.

PR 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0169315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.

PR 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PA Schlegel R, Deeds J, Berger A, Zhao X;



XX WPI; 2001-375006/39.  
DR new isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
PT  
PS Claim 1; Page 243; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful; to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 311 BP; 84 A; 67 C; 63 G; 94 T; 3 other;  
  
Query Match 4.3%; Score 86; DB 22; Length 311;  
Best Local Similarity 90.2%; Pred. No. 2.7e-08;  
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 1462  
DB 24 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 83  
  
QY 1463 GAGCTCAATAGGTACAAAGGAAGTACCTCTGTGACATCT 1504  
DB 84 GAGCTCAATAGGTACGGGGAAGTCACTCAGTTACTCT 125  
  
RESULT 14  
AAH70887  
ID AAH70887 standard; cDNA; 313 BP.  
XX  
AC AAH70887;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 2161.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
WPI; 2001-375006/39.  
XX  
PT New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
PS Claim 1; Page 456; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful; to assess if a patient is afflicted with

CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 313 BP; 83 A; 70 C; 63 G; 97 T; 0 other;  
  
Query Match 4.3%; Score 86; DB 22; Length 313;  
Best Local Similarity 90.2%; Pred. No. 2.7e-08;  
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 1462  
DB 2 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 61  
  
QY 1463 GAGCTCAATAGGTACAAAGGAAGTACCTCTGTGACATCT 1504  
DB 62 GAGCTCAATAGGTACGGGGAAGTCACTCAGTTACTCT 103  
  
RESULT 15  
AAH72445  
ID AAH72445 standard; cDNA; 299 BP.  
XX  
AC AAH72445;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 3719.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
WPI; 2001-375006/39.  
XX  
PT New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
PS Claim 1; Page 696; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful; to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 299 BP; 82 A; 64 C; 57 G; 96 T; 0 other;  
  
Query Match 4.2%; Score 84; DB 22; Length 299;  
Best Local Similarity 90.0%; Pred. No. 6.8e-08;  
Matches 90; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY	1405	ACTCAACACAGCTGCTTATGTTTATTATGGCACTTGTCATACAGAGACTGTTCCAGA	1464
Db	1	ACTCAACACAGCTGCTTATGTTTATTATGGCACTTGTCATACAGAGACTGTTCCAGA	60
OY	1465	GCTCAATAAGGTACAAAGGAGGTTACCTCTGTGCACATCT	1504
Db	61	GCTCAATAAGGTACGCGGGAAAGTCAACTCAGTACCTCT	100

Search completed: January 31, 2004, 10:41:08  
Job time : 553 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 18:21:00 ; Search time 708 Seconds  
(without alignments)  
10254.176 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: ttttgaatctccaaagt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO-NUC  
Gapop 60.0, Gapext 60.0

Searched: 2434938 seqs, 1822278265 residues

Word size: 6

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/prodata/2/pubpna/US09\_PUB.seq:  
14: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:  
17: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:  
18: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1824	91.6	3470	15	US-10-198-846-13016
2	1532	76.9	1716	13	US-10-117-722-922
3	1532	76.9	1716	13	US-10-037-270-922
4	918	46.1	1551	13	US-10-117-722-923
5	918	46.1	1551	13	US-10-037-270-923
6	444	22.3	535	11	US-09-918-995-22530
7	419	21.0	508	11	US-09-918-995-20205
8	397	19.9	397	11	US-09-918-995-5183
9	385	19.3	451	10	US-09-954-531-828
10	360	18.1	471	12	US-10-242-535A-25415
11	342	17.2	446	12	US-10-242-535A-49594
12	98	4.9	464	11	US-09-918-995-17035
13	60	3.0	60	13	US-09-908-975-12718
14	59	3.0	305	10	US-09-983-965-355
15	59	3.0	402	10	US-09-983-965-540

16	34	1.7	384	10	US-09-783-590-4845	Sequence 4845, Ap
17	32	1.6	280	10	US-09-783-590-4866	Sequence 4866, Ap
18	23	1.2	3673778	13	US-10-312-841-1	Sequence 1, Appli
c 19	22	1.1	756	13	US-10-027-632-164666	Sequence 164666,
c 20	22	1.1	756	14	US-10-027-632-164666	Sequence 164666,
c 21	22	1.1	3126	12	US-10-108-260A-777	Sequence 777, App
22	22	1.1	3448	13	US-10-027-632-115054	Sequence 115054,
23	22	1.1	3448	14	US-10-027-632-115054	Sequence 115054,
24	22	1.1	5807	13	US-10-311-455-1128	Sequence 1128, Ap
25	22	1.1	6009	13	US-10-311-455-802	Sequence 802, App
26	22	1.1	6070	13	US-10-311-455-1651	Sequence 1651, Ap
27	22	1.1	6070	13	US-10-240-485-131	Sequence 131, App
28	22	1.1	7049	13	US-10-311-455-130	Sequence 130, App
29	22	1.1	7049	13	US-10-240-452-6	Sequence 6, Appli
30	22	1.1	7238	13	US-10-311-455-422	Sequence 422, App
31	22	1.1	7771	13	US-10-311-455-1946	Sequence 1946, Ap
32	22	1.1	8576	13	US-10-311-455-2202	Sequence 2202, Ap
33	22	1.1	10716	13	US-10-311-455-1392	Sequence 1392, Ap
34	22	1.1	13511	13	US-10-311-455-253	Sequence 253, App
c 35	22	1.1	14798	13	US-10-311-455-1005	Sequence 1005, Ap
c 36	22	1.1	40862	13	US-10-311-455-2045	Sequence 2045, Ap
37	22	1.1	73334	13	US-10-311-455-2097	Sequence 2097, Ap
38	22	1.1	3673778	13	US-10-312-841-2	Sequence 2, Appli
c 39	21	1.1	729	13	US-10-027-632-155448	Sequence 155448,
c 40	21	1.1	729	14	US-10-027-632-155448	Sequence 155448,
c 41	21	1.1	737	9	US-09-563-817-372	Sequence 372, App
c 42	21	1.1	771	15	US-10-172-086-66	Sequence 66, Appl
c 43	21	1.1	910	13	US-10-190-312A-86	Sequence 86, Appl
c 44	21	1.1	1201	13	US-10-027-632-124069	Sequence 124069,
c 45	21	1.1	1201	14	US-10-027-632-124069	Sequence 124069,

ALIGNMENTS

RESULT 1  
US-10-198-846-13016  
; Sequence 13016, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10198.846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13016  
; LENGTH: 3470  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-846-13016

Query Match	91.6%	Score 1824;	DB 15;	Length 3470;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1824;	Conservative 0;	Mismatches 0;	Gaps 0;	
QY	147	GGACACACACAGAGGGGAGGAGTCACTTGGAAATCAGCAGATAGGTAGCAAAATTAAG	206	
Db	1647	GGACACACACAGAGGGGAGGAGTCACTTGGAAATCAGCAGATAGGTAGCAAAATTAAG	1706	
QY	207	GAGTATTCACAAAGTACCAATCGAGGAGCTATGTTGCCCTAATATGTTAGTCTGAAG	266	
Db	1707	GAGTATTCACAAAGTACCAATCGAGGAGCTATGTTGCCCTAATATGTTAGTCTGAAG	1766	
QY	267	GTGAAGATGATTTTATTGAGAGAGGTATGTTGTAATGGAAGATGATCTCCAGTGGAGG	326	

Db 1767 GTGAAGATGATTTATTTGAAGAGGATTTGTTGTAATGGAAGATGATTTCTCCAGTGGAGG 1826  
QY 327 CTGTGAGCACACCTAATACTCCCGAAACCTTGTCTGATGGAATAATTAGCAATTCATATG 386  
Db 1827 CTGTGAGCACACCTAATACTCCCGAAACCTTGTCTGATGGAATAATTAGCAATTCATATG 1886  
QY 387 TAGACTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGT 446  
Db 1887 TAGACTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGT 1946  
QY 447 TTTGTAATGATGTTGAAAGAAATGATAGAGAGAGTTGGACAGAGCTTGAACATTTGGT 506  
Db 1947 TTTGTAATGATGTTGAAAGAAATGATAGAGAGAGTTGGACAGAGCTTGAACATTTGGT 2006  
QY 507 CTGTCTATAGAGATATCTTGAATTTATGATCTATGATCTAATCAAACTAACAGATTTCAATG 566  
Db 2007 CTGTCTATAGAGATATCTTGAATTTATGATCTAATCAAACTAACAGATTTCAATG 2066  
QY 567 GTGCAATTTCCGTGATGCCAGCTTCTCTTAAGAGGATCAATTTGGCCCCCAAAATTTATGAAT 626  
Db 2067 GTGCAATTTCCGTGATGCCAGCTTCTCTTAAGAGGATCAATTTGGCCCCCAAAATTTATGAAT 2126  
QY 627 TCTTAAAGTCAAGAGGAGAGAGTTCCCAAGATATCTACAGAAACTTCTGCAGCATCCAG 686  
Db 2127 TCTTAAAGTCAAGAGGAGAGAGTTCCCAAGATATCTACAGAAACTTCTGCAGCATCCAG 2186  
QY 746 RAATCAGTAATAGTCAACTTCTTGGCAGACTTTCTTCCCTTAATTTGGGGGAAACACAAT 746  
Db 2187 RAATCAGTAATAGTCAACTTCTTGGCAGACTTTCTTCCCTTAATTTGGGGGAAACACAAT 2246  
QY 747 TTTCTGTAAGATATCTACAGATTAATCTTGGGAAATTTATAAATCTGTTCTCTGAA 806  
Db 2247 TTTCTGTAAGATATCTACAGATTAATCTTGGGAAATTTATAAATCTGTTCTCTGAA 2306  
QY 807 RAATTAATGAAGAGAGAGTCAATTTGGACCTTTTATCATCAATTTCAATTAATCTT 866  
Db 2307 RAATTAATGAAGAGAGAGTCAATTTGGACCTTTTATCATCAATTTCAATTAATCTT 2366  
QY 926 GTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACTTTCAG 926  
Db 2367 GTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACTTTCAG 2426  
QY 927 AAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACGGTCTGAAATA 986  
Db 2427 AAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACGGTCTGAAATA 2486  
QY 987 CAGAGAGAAAGCAAAATCAGATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATG 1046  
Db 2487 CAGAGAGAAAGCAAAATCAGATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATG 2546  
QY 1047 ATTACTGTGATGATCTAGAGGGGTAGTTTCCAGGTTTCTGACTGGCTTCCATCATCTCT 1106  
Db 2547 ATTACTGTGATGATGAGAGGGGTAGTTTCCAGGTTTCTGACTGGCTTCCATCATCTCT 2606  
QY 1107 TAATGGAACTCGAATCTCTTTAAAAACACCCCTGGAATGTATGATTAATCTTCT 1166  
Db 2607 TAATGGAACTCGAATCTCTTTAAAAACACCCCTGGAATGTATGATTAATCTTCT 2666  
QY 1167 AGTGTAACTAGAACAGCTATTCTAGGAGACCGTTTGGTCTCACTCAATAACATTTCTCA 1226  
Db 2667 AGTGTAACTAGAACAGCTATTCTAGGAGACCGTTTGGTCTCACTCAATAACATTTCTCA 2726  
QY 1227 GAGATGCTATTTCTGTGAAACACTGACCTGCTCTCTCCAGATAGCAAAAAGGAG 1286  
Db 2727 GAGATGCTATTTCTGTGAAACACTGACCTGCTCTCTCCAGATAGCAAAAAGGAG 2786  
QY 1287 CAAAACAGACTTTTGAAGAAATGATGATTTACATTTCCAGATCTGTTAGTCAAGTGTATTG 1346  
Db 2787 CAAAACAGACTTTTGAAGAAATGATGATTTACATTTCCAGATCTGTTAGTCAAGTGTATTG 2846  
QY 1347 GTGAAGAAACCAAGTATGAAGCATCAGACTTTCTGTTTGTGGCTTACAGCAACCAAGTAC 1406

Db 2847 GTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGTGGCTTACAGCAACAGTAC 2906  
QY 1407 TCAACAAGCAGCTGACTTATGTTTTATTGGACATTTGTGATACAGAACTGTTTCCAGAGC 1466  
Db 2907 TCAACAAGCAGCTGACTTATGTTTTATTGGACATTTGTGATACAGAACTGTTTCCAGAGC 2966  
QY 1467 TCAATAAGGTACAAAGGAAGTTACTCTCTGTGACATCTTGGATGTAAACACTTGGATTG 1526  
Db 2967 TCAATAAGGTACAAAGGAAGTTACTCTCTGTGACATCTTGGATGTAAACACTTGGATTG 3026  
QY 1527 GTATAGATTAACCCATTGAAATTTCTGCTGTGCGAGGGTGTAGAAATTTACTTTTTTGG 1586  
Db 3027 GTATAGATTAACCCATTGAAATTTCTGCTGTGCGAGGGTGTAGAAATTTACTTTTTTGG 3086  
QY 1587 GTATATTTCTTATATATATTTATGATCGCTGTCTGAAATTTTACTTATTTTCTTTTTT 1646  
Db 3087 GTATATTTCTTATATATATTTATGATCGCTGTCTGAAATTTTACTTATTTTCTTTTTT 3146  
QY 1647 AATAAGACTAACACAAACTTAATGATTTAAAGTGAATTTGAGTCTCATAGTCTTTTCATTG 1706  
Db 3147 AATAAGACTAACACAAACTTAATGATTTAAAGTGAATTTGAGTCTCATAGTCTTTTCATTG 3206  
QY 1707 CTAGCTGTGATCCAAATTTTATAGAACATAAGTCACTTGTATTGCCATTTTAAAGA 1766  
Db 3207 CTAGCTGTGATCCAAATTTTATAGAACATAAGTCACTTGTATTGCCATTTTAAAGA 3266  
QY 1767 GAAATTTCTATATGATTTATGGCAACAGATAGACTGTAAACTTCGTATTGTATAGC 1826  
Db 3267 GAAATTTCTATATGATTTATGGCAACAGATAGACTGTAAACTTCGTATTGTATAGC 3326  
QY 1827 TTTGAAATTAATTTATGCTAGTATGAGAAACAGAACTAAAGATCTGATTTCTTAGAGTT 1886  
Db 3327 TTTGAAATTAATTTATGCTAGTATGAGAAACAGAACTAAAGATCTGATTTCTTAGAGTT 3386  
QY 1887 AATATATTTTAGTAGATTGGTTTTCTTTTATTTTATTTTGTACATAGTTAACTGTGTATC 1946  
Db 3387 AATATATTTTAGTAGATTGGTTTTCTTTTATTTTATTTTGTACATAGTTAACTGTGTATC 3446  
QY 1947 TATAATAAGCATCTCATATGAG 1970  
Db 3447 TATAATAAGCATCTCATATGAG 3470

## RESULT 2

US-10-117-722-922  
; Sequence 922, Application US/10117722  
; Publication No. US20030219744A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2BCIP  
; CURRENT APPLICATION NUMBER: US/10/117,722  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1543)

US-10-117-722-922

Query Match		76.9%	Score 1532;	DB 13;	Length 1716;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 1532;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	147	CGAACACACAGAAAAGGGGAGAGATCATTTGGAAATCAGCAGAGATAGGTAGCAAAATTAAG	206		
DB	177	GGACACACAGAAAAGGGGAGAGATCATTTGGAAATCAGCAGAGATAGGTAGCAAAATTAAG	236		
QY	207	GAGTATTTCAAAAGTACCAATGAGGAGGAGCTATGTCCTTAATPATGGTGTAGCTGAAG	266		
DB	237	GAGTATTTCAAAAGTACCAATGAGGAGGAGCTATGTCCTTAATPATGGTGTAGCTGAAG	296		
QY	267	GTGAGATGATTTATTTAGGAAGAGTATTTGTTGTAATGGAAGATGATTTCTCCAGTGGAGG	326		
DB	297	GTGAGATGATTTATTTAGGAAGAGTATTTGTTGTAATGGAAGATGATTTCTCCAGTGGAGG	356		
QY	327	CTGTGAGCACACTAATCTCCCGAAACCTTCTGTCATGGAATAGCAATTCATATG	386		
DB	357	CTGTGAGCACACTAATCTCCCGAAACCTTCTGTCATGGAATAGCAATTCATATG	416		
QY	387	TAGACTTTTTCGAGGATCCCTCTCTGAAAGAGAGAGAAAAGAAAGAAATTCCTGTGT	446		
DB	417	TAGACTTTTTCGAGGATCCCTCTCTGAAAGAGAGAGAAAAGAAAGAAATTCCTGTGT	476		
QY	447	TTTGTATTCATGTTGAAAGAAATGATAGAGAGAGAGTTGGACACGAGCTGAACATGGT	506		
DB	477	TTTGTATTCATGTTGAAAGAAATGATAGAGAGAGAGTTGGACACGAGCTGAACATGGT	536		
QY	507	CTGTCTATAGAGATATCTTGAATCTATGTACTTGAATCAAACTACAGAAATTCATG	566		
DB	537	CTGTCTATAGAGATATCTTGAATCTATGTACTTGAATCAAACTACAGAAATTCATG	596		
QY	567	GTGCATTTCTGATGCCAGCTTCCCTCTAAGAGGATCATTCGGCCCAAAATTAATGAAT	626		
DB	597	GTGCATTTCTGATGCCAGCTTCCCTCTAAGAGGATCATTCGGCCCAAAATTAATGAAT	656		
QY	627	TCCTAAAGTCAAGAGGAGAGAGTTCCAAAGATATCTCAGAAACTTCTGCAGATCCAG	686		
DB	657	TCCTAAAGTCAAGAGGAGAGAGTTCCAAAGATATCTCAGAAACTTCTGCAGATCCAG	716		
QY	687	AACAGATTAATGATCTGAGCTTCTGAGAGCTTCTTCCCTTAATGGTGGGAAACACAAAT	746		
DB	717	AACAGATTAATGATCTGAGCTTCTGAGAGCTTCTTCCCTTAATGGTGGGAAACACAAAT	776		
QY	747	TTCTGATTAAGTACTACAGATGTAAATCTTGGGAAATTAATAATCTGTTCTCTGAA	806		
DB	777	TTCTGATTAAGTACTACAGATGTAAATCTTGGGAAATTAATAATCTGTTCTCTGAA	836		
QY	807	AACATTAAGAGAGAGAGAGTTCAGACTTTGGAACCTTTTATCATGAATTTCAATTAATCTT	866		
DB	837	AACATTAAGAGAGAGAGAGTTCAGACTTTGGAACCTTTTATCATGAATTTCAATTAATCTT	896		
QY	867	GTGAGTCTCCAAAGCCCTAAACCAAGTAGCAGAACTGACCAATCTCAGCCCTACTTCAG	926		
DB	897	GTGAGTCTCCAAAGCCCTAAACCAAGTAGCAGAACTGACCAATCTCAGCCCTACTTCAG	956		
QY	927	AAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAATAATTCGAAACCGTCTGAAATA	986		
DB	957	AAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAATAATTCGAAACCGTCTGAAATA	1016		
QY	987	CAGAGAGAGCAAAATCAGAAATTTATTTGAGGAGTATGATCTGTAGAGAGAGCTATG	1046		
DB	1017	CAGAGAGAGCAAAATCAGAAATTTATTTGAGGAGTATGATCTGTAGAGAGAGCTATG	1076		
QY	1047	ATTACCTGATGATGAGAGCGGTAGTTTTCAGAGTCTCTGACTGGCTTCATCATCTCT	1106		
DB	1077	ATTACCTGATGATGAGAGCGGTAGTTTTCAGAGTCTCTGACTGGCTTCATCATCTCT	1136		
QY	1107	TAATGGGAATCGAATCTCTTTTAAACACACCTCGAAATGATATGATTAATCTATCTTC	1166		
DB	1137	TAATGGGAATCGAATCTCTTTTAAACACACCTCGAAATGATATGATTAATCTATCTTC	1196		

QY	1167	AGTGTAACCTAGAACAGCTATTTTCAGGAGCAGCGTTTGGTCTCACTATCAACACTTCTCA	1226
DB	1197	AGTGTAACCTAGAACAGCTATTTTCAGGAGCAGCGTTTGGTCTCACTATCAACACTTCTCA	1256
QY	1227	GAGATGCTATATTTCTGTGAAAAACACATGAAACCTGCGTCTCTCCAGATATAAGCAAAAGGAG	1286
DB	1257	GAGATGCTATATTTCTGTGAAAAACACATGAAACCTGCGTCTCTCCAGATATAAGCAAAAGGAG	1316
QY	1287	CAAAACAGACTTTTGGAGAAATGATTAATTAATCCAGATCTGTTAGTCAAGTGTATTG	1346
DB	1317	CAAAACAGACTTTTGGAGAAATGATTAATTAATCCAGATCTGTTAGTCAAGTGTATTG	1376
QY	1347	GTCAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGCGCTTACAGCAACCAAGTAC	1406
DB	1377	GTCAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGCGCTTACAGCAACCAAGTAC	1436
QY	1407	TCACCAAGCAGCTGACTTATGTTTATTTGACATTTGTGATACAGGAACCTGTTTCCAGAGC	1466
DB	1437	TCACCAAGCAGCTGACTTATGTTTATTTGACATTTGTGATACAGGAACCTGTTTCCAGAGC	1496
QY	1467	TCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAAACACTTGGATTG	1526
DB	1497	TCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAAACACTTGGATTG	1556
QY	1527	GTATAGATAACCCATTTGAAATTTCTGCTGTGCGAGGCTGTAGAAATTTACTTTTTTGG	1586
DB	1557	GTATAGATAACCCATTTGAAATTTCTGCTGTGCGAGGCTGTAGAAATTTACTTTTTTGG	1616
QY	1587	GTATATTTCTTATATATATATGATGACATCGCTGTCTGAAATTTTAGTTATTTTGTTTTT	1646
DB	1617	GTATATTTCTTATATATATATGATGACATCGCTGTCTGAAATTTTAGTTATTTTGTTTTT	1676
QY	1647	AATAAGACTTAACACAACTTAATGATTAATAA 1678	
DB	1677	AATAAGACTTAACACAACTTAATGATTAATAA 1708	

RESULT 3

US-10-037-270-922  
; Sequence 922, Application US/10037270  
; Publication No. US20030104529A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feliyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Tillinghast, John  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/10/037,270  
; PRIOR APPLICATION NUMBER: 2002-01-04  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pc\_fl\_genes version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716

```
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (71)..(1543)
US-10-037-270-922

Query Match      76.9%  Score 1532; DB 15; Length 1716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGAACACACAGAAAGGGGAGAAATCATTTTGAATCAGCAGAAATAGGTAGCAAAATTAAG 206
DB 177 GGAACACACAGAAAGGGGAGAAATCATTTTGAATCAGCAGAAATAGGTAGCAAAATTAAG 236
QY 207 GAGTATTCAAAAGTACCAATAGGAGGAGTATGTTGGCTTAATATGTTGGTATAGTGAAG 266
DB 237 GAGTATTCAAAAGTACCAATAGGAGGAGTATGTTGGCTTAATATGTTGGTATAGTGAAG 296
QY 267 GTGAAGATGATTTATTTGAAGAGGTATGTTTGAATGGAAGATGATCTCCAGTGGAGG 326
DB 297 GTGAAGATGATTTATTTGAAGAGGTATGTTTGAATGGAAGATGATCTCCAGTGGAGG 356
QY 327 CTGTGAGCACACCTAATCTCCCGAAACCTTGCTGATGAAATTTAGCATTCATATG 386
DB 357 CTGTGAGCACACCTAATCTCCCGAAACCTTGCTGATGAAATTTAGCATTCATATG 416
QY 387 TAGACTTTTGTAGGATCCCTCTGAAAGGAAGGAGAAAGAAAGAAATTCCTGTGT 446
DB 417 TAGACTTTTGTAGGATCCCTCTGAAAGGAAGGAGAAAGAAAGAAATTCCTGTGT 476
QY 447 TTTGATTTGATTTGAAAGAAATGATGAGAGCAGATTTGGACAGAGCTGAAACATTTGGT 506
DB 477 TTTGATTTGATTTGAAAGAAATGATGAGAGCAGATTTGGACAGAGCTGAAACATTTGGT 536
QY 507 CTGTCTATAGAGATATCTGAAATCTATGTTGATCAAACTAACAGAAATTTTCATG 566
DB 537 CTGTCTATAGAGATATCTGAAATCTATGTTGATCAAACTAACAGAAATTTTCATG 596
QY 567 GTGCATTTTCTGATCCCGAGCTTCTTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAAT 626
DB 597 GTGCATTTTCTGATCCCGAGCTTCTTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAAT 656
QY 627 TCTTAAAGTCAAAGGAGGAGATGTTCCAGAAATATCTACAGAAATCTTGAGAGATCCAG 686
DB 657 TCTTAAAGTCAAAGGAGGAGATGTTCCAGAAATATCTACAGAAATCTTGAGAGATCCAG 716
QY 687 AACTGAGTAATAGTCAACTTCTGGCAGACTTTCTTCCCTTAATGGTGGGAAACACAAT 746
DB 717 AACTGAGTAATAGTCAACTTCTGGCAGACTTTCTTCCCTTAATGGTGGGAAACACAAT 776
QY 747 TCTTGTATAGATATCTACAGATGTAATCTTGGGAAATATATAAATCTCTCTCGAA 806
DB 777 TCTTGTATAGATATCTACAGATGTAATCTTGGGAAATATATAAATCTCTCTCGAA 836
QY 807 AACTAATGAAGAGAAAGGTACAGACTTTTGGAACTTTTATCATGAATTTCAATTAATCTT 866
DB 837 AACTAATGAAGAGAAAGGTACAGACTTTTGGAACTTTTATCATGAATTTCAATTAATCTT 896
QY 867 GTGAGTCTCAAAGCCTTAAACCAAGTAGACAGAACTGACCAATTTCTAGCCCTACTTCAG 926
DB 897 GTGAGTCTCAAAGCCTTAAACCAAGTAGACAGAACTGACCAATTTCTAGCCCTACTTCAG 956
QY 927 AAAACAACAGAGCTTTTCAATGATCTGTTTAAAATTAATGCAACCGTCTGAAATA 986
DB 957 AAAACAACAGAGCTTTTCAATGATCTGTTTAAAATTAATGCAACCGTCTGAAATA 1016
QY 987 CAGAGAGAAAGCAAAATCAGAAATTTTATGAGGTGATGACTGTAGAAGAGTCTATG 1046
DB 1017 CAGAGAGAAAGCAAAATCAGAAATTTTATGAGGTGATGACTGTAGAAGAGTCTATG 1076
QY 1047 ATTACTGATGATGTAGACCGGTGTTTCCAGGTTCTGACTGGCTTCATCATCTCT 1106
```

```
DB 1077 ATTACTGATGATGTAGGACGGGTAGTTTCCAGGTTCTCTGACTGGCTTCATCATCTCT 1136
QY 1107 TAATGGGAACCTCGAATCCTCTTTTAAACACACCTCGGAATATGATGATGATGATCTTTC 1166
DB 1137 TAATGGGAACCTCGAATCCTCTTTTAAACACACCTCGGAATATGATGATGATGATCTTTC 1196
QY 1167 AGTGTAAACTAGAACAGACTATTTTCAAGGACACCGTTTGGTCTCACTCATACACTTCTCA 1226
DB 1197 AGTGTAAACTAGAACAGACTATTTTCAAGGACACCGTTTGGTCTCACTCATACACTTCTCA 1256
QY 1227 GAGATGCTATATTTGTGAAACACACTGAACTCGCTCTCTCCAAAGATAAGCAAAAGGAG 1286
DB 1257 GAGATGCTATATTTGTGAAACACACTGAACTCGCTCTCTCCAAAGATAAGCAAAAGGAG 1316
QY 1287 CAAAACAGACTTTTGAAGAAATGATGAATTAACATCCAGATCTGTAGTCAAGTGTATTG 1346
DB 1317 CAAAACAGACTTTTGAAGAAATGATGAATTAACATCCAGATCTGTAGTCAAGTGTATTG 1376
QY 1347 GTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTGTGATGGCTTACAGCAACCAAGTAC 1406
DB 1377 GTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTGTGATGGCTTACAGCAACCAAGTAC 1436
QY 1407 TCACAAGCAGCTGACTTATGTTTATTTGACATTTGTGATACAGGAATCTGTTTCCAGAGC 1466
DB 1437 TCACAAGCAGCTGACTTATGTTTATTTGGACATTTGTGATACAGGAATCTGTTTCCAGAGC 1496
QY 1467 TCAATAAGGTACAAAGGAAGTTACTCTGTGACATCTTTGGATGTAACACACTTGGATTTG 1526
DB 1497 TCAATAAGGTACAAAGGAAGTTACTCTGTGACATCTTTGGATGTAACACACTTGGATTTG 1556
QY 1527 GTATAGAAATACCCATTTGAATTTCTGCTGTGGAGGGTGTAGAAATTTACTTTTTTTGG 1586
DB 1557 GTATAGAAATACCCATTTGAATTTCTGCTGTGGAGGGTGTAGAAATTTACTTTTTTTGG 1616
QY 1587 GTATATTTCTTATATATATATGATGATGCTGTCTGAAATTTTACTTTTTTTGTTTTT 1646
DB 1617 GTATATTTCTTATATATATATGATGATGCTGTCTGAAATTTTACTTTTTTTGTTTTT 1676
QY 1647 AATAAGACTAACACAAACTTAATGATTAATA 1678
DB 1677 AATAAGACTAACACAAACTTAATGATTAATA 1708

RESULT 4
US-10-117-722-923
; Sequence 923, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1: Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 923
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(1378)
```

US-10-117-722-923

Query Match 46.1%; Score 918; DB 13; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGACACACAGAAAGGGAGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATTTAAAG 206  
DB 177 GGACACACAGAAAGGGAGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATTTAAAG 236

QY 207 GAGTATTCAGAAAGTACCAATGAGGAGCTATGTTGCTTAATATGTTAGTGTAGCTGAAG 266  
DB 237 GAGTATTCAGAAAGTACCAATGAGGAGCTATGTTGCTTAATATGTTAGTGTAGCTGAAG 296

QY 267 GTGAGATGATTTTATTTAGAGAGGATTTGTTGTAATGGAAGATGATTCCTCAGTGGAGG 326  
DB 297 GTGAGATGATTTTATTTAGAGAGGATTTGTTGTAATGGAAGATGATTCCTCAGTGGAGG 356

QY 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATCGAAATTTAGCATTCATATG 386  
DB 357 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATCGAAATTTAGCATTCATATG 416

QY 387 TAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTGT 446  
DB 417 TAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTGT 476

QY 447 TTTGTATTTGATTTGAAAGAAATGATAGAGAGCAGTGTGACACGAGCCTGGAACATGGT 506  
DB 477 TTTGTATTTGATTTGAAAGAAATGATAGAGAGCAGTGTGACACGAGCCTGGAACATGGT 536

QY 507 CTGTCTATAGAGATATCTTGAATTTCTATGTTGATCTGATCAAACTAACAGAAATTTTCATG 566  
DB 537 CTGTCTATAGAGATATCTTGAATTTCTATGTTGATCTGATCAAACTAACAGAAATTTTCATG 596

QY 567 GTGCATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCGCAAAATTTATGAAT 626  
DB 597 GTGCATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCGCAAAATTTATGAAT 656

QY 627 TCTTAAAGTCAAGAGGAGAGAGTTCCAGAAATATCTACAGAACTTCTGACAGATCCAG 686  
DB 657 TCTTAAAGTCAAGAGGAGAGAGTTCCAGAAATATCTACAGAACTTCTGACAGATCCAG 716

QY 687 AACTGATATAGTCAACTTCTGACAGACTTCTTCTCCCTTAATGTTGGGGAACACCAAT 746  
DB 717 AACTGATATAGTCAACTTCTGACAGACTTCTTCTCCCTTAATGTTGGGGAACACCAAT 776

QY 747 TCTTGAATAGATATCAAGATGTAATCTTGGGAAATTTATAAAATCTGTTCTCTGGAA 806  
DB 777 TCTTGAATAGATATCAAGATGTAATCTTGGGAAATTTATAAAATCTGTTCTCTGGAA 836

QY 807 AACTTAATGAAGAGAGAGTCCAGATTTGGAACCTTTTATCATGAATTTTCAATTAATCTT 866  
DB 837 AACTTAATGAAGAGAGAGTCCAGATTTGGAACCTTTTATCATGAATTTTCAATTAATCTT 896

QY 867 GTGAGTCTCAGAGCTTAACCAAGTACAGAGACTGACCATTTCTCAGCCCTTCTCAG 926  
DB 897 GTGAGTCTCAGAGCTTAACCAAGTACAGAGACTGACCATTTCTCAGCCCTTCTCAG 956

QY 927 AARACACAGAGCTTTCAATGATCTGTTTAAATATGCAAAACCGTCTGGAATA 986  
DB 957 AARACACAGAGCTTTCAATGATCTGTTTAAATATGCAAAACCGTCTGGAATA 1016

QY 987 CAGAGAGAGCAAAATCAGAAATTTATTTATGAGGATGATGATGTAGAGAGGATCTATG 1046  
DB 1017 CAGAGAGAGCAAAATCAGAAATTTATTTATGAGGATGATGATGTAGAGAGGATCTATG 1076

QY 1047 ATTACCTGATGTATGTAG 1064  
DB 1077 ATTACCTGATGTATGTAG 1094

RESULT 5  
US-10-037-270-923

Sequence 923, Application US/10037270  
Publication No. US20030104529A1

GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunqing  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillinghast, John  
APPLICANT: Dmanac, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt FL\_genes Version 1.0  
SEQ ID NO 923  
LENGTH: 1551  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (71)...(1378)  
US-10-037-270-923

Query Match 46.1%; Score 918; DB 15; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGACACACAGAAAGGGAGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATTTAAAG 206  
DB 177 GGACACACAGAAAGGGAGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATTTAAAG 236

QY 207 GAGTATTCAGAAAGTACCAATGAGGAGCTATGTTGCTTAATATGTTAGTGTAGCTGAAG 266  
DB 237 GAGTATTCAGAAAGTACCAATGAGGAGCTATGTTGCTTAATATGTTAGTGTAGCTGAAG 296

QY 267 GTGAGATGATTTTATTTAGAGAGGATTTGTTGTAATGGAAGATGATTCCTCAGTGGAGG 326  
DB 297 GTGAGATGATTTTATTTAGAGAGGATTTGTTGTAATGGAAGATGATTCCTCAGTGGAGG 356

QY 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATCGAAATTTAGCATTCATATG 386  
DB 357 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATCGAAATTTAGCATTCATATG 416

QY 387 TAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTTCTGTGT 446  
DB 417 TAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTTCTGTGT 476

QY 447 TTTGTATTTGATTTGAAAGAAATGATAGAGAGCAGTGTGACAGCCTGGAACATTTGTT 506  
DB 477 TTTGTATTTGATTTGAAAGAAATGATAGAGAGCAGTGTGACAGCCTGGAACATTTGTT 536

QY 507 CTGTCTATAGAGATATCTTGAATTTCTATGTTGATCTTGAATCAAACTAACAGAAATTTTCATG 566  
DB 537 CTGTCTATAGAGATATCTTGAATTTCTATGTTGATCTTGAATCAAACTAACAGAAATTTTCATG 596

QY 567 GTGCATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCGCAAAATTTATGAAT 626



```
Db 597 GTGCAATTCCTGATGCCAGCTTCTCTTAGAGGATCAATGGCCCCAATAATATGAT 656
Qy 627 TCTTAAAGTCAAGAGGAGAGTCCAGAAATATCTACAGAACTTCTGAGCATCCAG 686
Db 657 TCTTAAAGTCAAGAGGAGAGTCCAGAAATATCTACAGAACTTCTGAGCATCCAG 716
Qy 687 AACTGAGTAATAGTCACTCTGCGAGACTTCTTCCCTTAAGTGGGGGAAACAAAT 746
Db 717 AACTGAGTAATAGTCACTCTGCGAGACTTCTTCCCTTAAGTGGGGGAAACAAAT 776
Qy 747 TCTTGTAAAGTACTACAGATGTAATCTTGGGAAAATATATAAAATCTGTCCTGGAA 806
Db 777 TCTTGTAAAGTACTACAGATGTAATCTTGGGAAAATATATAAAATCTGTCCTGGAA 836
Qy 807 AACTAATGAAGAGAAAGGTGAGATTTGGAACCTTTTATCATGAATTCATTAATCTT 866
Db 837 AACTAATGAAGAGAAAGGTGAGATTTGGAACCTTTTATCATGAATTCATTAATCTT 896
Qy 867 GTGAGTCTCAAAGCTTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 926
Db 897 GTGAGTCTCAAAGCTTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 956
Qy 927 AAACAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAAATA 986
Db 957 AAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAAATA 1016
Qy 987 CAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGGATCTATG 1046
Db 1017 CAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGGATCTATG 1076
Qy 1047 ATTACTGATGATGATG 1064
Db 1077 ATTACTGATGATGATG 1094
```

## RESULT 6

```
US-09-918-995-22530
; Sequence 22530, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22530
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(535)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22530
```

```
Query Match 22.3%; Score 444; DB 11; Length 535;
Best Local Similarity 100.0%; Pred. No. 5.5e-206;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1393 ACAGCAACAGTACTCAACAGAGAGTCACTTATGTTTATGACATTTGATACAGGA 1452
Db 34 ACAGCAACAGTACTCAACAGAGAGTCACTTATGTTTATGACATTTGATACAGGA 93
Qy 1453 ACTGTTTCCAGAGCTCAATAAGGTACAAAGAGTACCTCTGACATCTTGATGTA 1512
Db 94 ACTGTTTCCAGAGCTCAATAAGGTACAAAGAGTACCTCTGACATCTTGATGTA 153
```

```
Qy 1513 AACACTTGGATTGGTATAGATAACCCATTGAAATTTCTGCTGTGCGAGGTTGTTAGAA 1572
Db 154 AACACTTGGATTGGTATAGATAACCCATTGAAATTTCTGCTGTGCGAGGTTGTTAGAA 213
Qy 1573 ATTACTTTTTTGGGTATATTTCTTATATATATATATATATATATATATATATATAT 1632
Db 214 ATTACTTTTTTGGGTATATTTCTTATATATATATATATATATATATATATATATAT 273
Qy 1633 TATTTTTTTGTTTTTAAATAGACTAACACAACTTAAATGATTAAGATGATTGAGTCTCA 1692
Db 274 TATTTTTTTGTTTTTAAATAGACTAACACAACTTAAATGATTAAGATGATTGAGTCTCA 333
Qy 1693 TAGTCTTTTCAATTTGCTAGCTGTGATCCAAATTTTATTTAGAACATAAAGTCACTTGTATTG 1752
Db 334 TAGTCTTTTCAATTTGCTAGCTGTGATCCAAATTTTATTTAGAACATAAAGTCACTTGTATTG 393
Qy 1753 CCATTTTTAAAGAGAAAATTCATATGATGTTATGCGCAACACAGATAAGACTGATAAAT 1812
Db 394 CCATTTTTAAAGAGAAAATTCATATGATGTTATGCGCAACACAGATAAGACTGATAAAT 453
Qy 1813 TCGTATTGTATAGCTTTGAAAATA 1836
Db 454 TCGTATTGTATAGCTTTGAAAATA 477
```

## RESULT 7

```
US-09-918-995-20205
; Sequence 20205, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20205
; LENGTH: 508
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-20205
```

```
Query Match 21.0%; Score 419; DB 11; Length 508;
Best Local Similarity 100.0%; Pred. No. 8.9e-194;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1190 CAGGAGCACCGTTTGGTCTCACTATAACACTTCTCAGAGATGCTTATATCTGTGAAAAC 1249
Db 65 CAGGAGCACCGTTTGGTCTCACTATAACACTTCTCAGAGATGCTTATATCTGTGAAAAC 124
Qy 1250 ACTGAACCTCGCTCTCTCAGAGATAAGCAAAAAGAGCAAAAACAGACTTTTGAAGAAATG 1309
Db 125 ACTGAACCTCGCTCTCTCAGAGATAAGCAAAAAGAGCAAAAACAGACTTTTGAAGAAATG 184
Qy 1310 ATGAATTACATTCAGATCTGTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATGAAAGC 1369
Db 185 ATGAATTACATTCAGATCTGTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATGAAAGC 244
Qy 1370 ATCAGACTTCTGTTTGTAGGCTTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 1429
Db 245 ATCAGACTTCTGTTTGTAGGCTTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 304
Qy 1430 TTATTGGACATTTGATACAGAACTGTTTTCAGAGCTCAATAAGGTACAAAGAAAGTT 1489
Db 305 TTATTGGACATTTGATACAGAACTGTTTTCAGAGCTCAATAAGGTACAAAGAAAGTT 364
Qy 1490 ACTCTGTGACATCTTGGATGTAACACTTGGATTTGGTATAGATATACCCATTCGAAT 1549
Db 365 ACTCTGTGACATCTTGGATGTAACACTTGGATTTGGTATAGATATACCCATTCGAAT 424
```



```

; DESCRIPT: test
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(464)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-17035

Query Match      4.9%      Score 98:  DB 11:  Length 464;
Best Local Similarity 100.0%;  Pred. No. 5.4e-37;
Matches 98;  Conservative 0;  Mismatches 0;  Indels

OV      1887  AACTATATTTTACGATGGTTTTCCCTTTTTTTTTTTTATTTGTCACATAGTTTA

```

Db 187 AATATATTTAGTAGATTGGTTTCTCTTTTATTTTGTACATAGTTAACTGTGTATC 246  
Qy 1947 TATAATAAGCATCTATATGATGTTTAAATAATAA 1984  
Db 247 TATAATAAGCATCTATATGATGTTTAAATAATAA 284

RESULT 13  
US-09-908-975-12718  
; Sequence 12718, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, AVI  
; APPLICANT: WASSERMAN, ALON  
; APPLICANT: MINTZ, ELI  
; APPLICANT: FAIGLER, SIMCHON  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE  
; FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/287,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 12718  
; LENGTH: 60  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-975-12718

Query Match 3.0%; Score 60; DB 13; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.6e-18;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 891 GTAGACAGACTGACCATCTCAGCCCTACTTCAGAAACACACAGAGCTTTCAATG 950  
Db 1 GTAGACAGACTGACCATCTCAGCCCTACTTCAGAAACACACAGAGCTTTCAATG 60

RESULT 14  
US-09-983-965-355  
; Sequence 355, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 355  
; LENGTH: 305  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 42-BOVMS1-019-Q1-E1-C6  
US-09-983-965-355

Query Match 3.0%; Score 59; DB 10; Length 305;  
Best Local Similarity 100.0%; Pred. No. 5.8e-18;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 770 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAGAGAAAGTCA 828  
Db 118 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAGAGAAAGTCA 176

RESULT 15  
US-09-983-965-540  
; Sequence 540, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 540  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (339)  
; OTHER INFORMATION:  
; OTHER INFORMATION: Clone ID: 64-BOVMS1-017-Q1-E1-H12  
US-09-983-965-540

Query Match 3.0%; Score 59; DB 10; Length 402;  
Best Local Similarity 100.0%; Pred. No. 5.9e-18;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 770 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAGAGAAAGTCA 828  
Db 45 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAGAGAAAGTCA 103

Search completed: January 31, 2004, 21:44:01  
Job time : 716 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 05:18:46 : Search time 5257 Seconds  
(without alignments)  
3618.599 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 465

Sequence: 1 WLIHFCLTRVTKRGESF.....ELFPELNKVKQVTSVTSWM 465

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Egapop 6.0, Egapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2 1/USPTO.spool/US09744313/runat 30012004 113612 4590/app query.fasta\_1.647  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09744313 @CGN 1 1 3508 @runat 30012004 113612 4590 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pi.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	465	100.0	1992	6	AX054819 Sequence
2	455	97.8	2661	9	AY044865 Homo sapi
3	455	97.8	3038	9	BC005110 Homo sapi
4	390	83.9	2925	9	BC048520 Homo sapi
5	374	80.4	1593	9	AF121863 Homo sapi
6	368	79.1	3576	9	AK026479 Homo sapi
7	330	71.0	3145	6	AX512835 Sequence
8	330	71.0	3145	6	AK000362 Homo sapi
9	269	57.8	3616	9	AK095380 Homo sapi
10	163	35.1	968	9	BSA420581
11	141	30.3	425	6	BD113855 EST and e
12	119	25.6	358	6	BD026500 Sequence
13	109	23.4	1782	10	BC043328 Mus muscu
14	67	14.4	123010	9	AL589666 Human DNA
15	67	14.4	278898	2	AC118713
16	59	12.7	224551	2	AC111832 Rattus no
17	55	11.8	278375	2	AC130093 Rattus no
18	47	10.1	454	11	G30543 human STS S
19	25	5.4	311	6	AX185131 Sequence
20	25	5.4	313	6	AX186466 Sequence
21	24	5.2	299	6	AX188027 Sequence
22	21	4.5	183972	2	EX537259 Danio rer
23	12	2.6	36154	2	AC100446 Mus muscu
24	11	2.4	160411	2	AC135935 Rattus no
25	11	2.4	200417	2	AC134192 Rattus no
26	11	2.4	230404	2	AC097160 Rattus no
27	10	2.2	214222	2	AC074160 Mus muscu
28	9	1.9	188	6	BD123051 EST and e
29	9	1.9	347	6	BD123053 EST and e
30	9	1.9	367	6	BD123050 EST and e
31	9	1.9	377	6	BD123049 EST and e
32	9	1.9	506	6	BD123048 EST and e
33	9	1.9	543	6	BD123052 EST and e
34	9	1.9	684	8	AF513658 Aphanoal
35	9	1.9	815	5	AF390220 Sarda sar
36	9	1.9	821	5	AF390239 Sarda sar
37	9	1.9	824	5	AF390246 Sarda sar
38	9	1.9	830	5	AF390225 Sarda sar
39	9	1.9	839	5	AF390245 Sarda sar
40	9	1.9	845	5	AF390232 Sarda sar
41	9	1.9	846	5	AF390235 Sarda sar
42	9	1.9	847	5	AF390212 Sarda sar
43	9	1.9	847	5	AF390216 Sarda sar
44	9	1.9	847	5	AF390217 Sarda sar
45	9	1.9	847	5	AF390217 Sarda sar

ALIGNMENTS

RESULT 1

AX054819  
 LOCUS AX054819 1922 bp DNA linear PAT 13-JAN-2001  
 DEFINITION Sequence 3 from Patent WO0073334.  
 ACCESSION AX054819  
 VERSION AX054819.1 GI:12228268  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Yue H., Tang, Y. T. and Azimzai, Y.  
 TITLE Human sorting nexin  
 JOURNAL Patent: WO 00/73334-A 3 07-DEC-2000;  
 INCYTE Genomics, Inc. (US)  
 FEATURES  
 Location/Qualifiers  
 1..1992  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 671 a 308 c 372 g 641 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 1992  
 Score: 465.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-744-313A-1 (1-465) x AX054819 (1-1992)  
 QY 1 MetTyrLeuIleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyValSerPhe 20  
 DB 116 AUGTACTTGATCACTTTTGGTTTATATTCAGAGACACACAGAAAGGGGAGATCATTT 175  
 QY 21 GlyIleSerArgIleGlySerIleLysGlyValPheLysSerThrThrMetGluGly 40  
 DB 176 GGAATCAGCAGATAGTAGCAAAATTAAGGAGTATTCAAAGTACCACCAATGGAGGA 235  
 QY 41 AlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGlyIle 60  
 DB 236 GCTATCTGCCTAATATATGCTAGCTGAAGTGAAGATGATTTTATTGAAGAGGTATT 295  
 QY 61 ValValMetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80  
 DB 296 GTTGTAATGGAGATGATTTCTCAGTGGAGGCTGTGAGACACACCTTAATCTCCCGAATC 355  
 QY 81 LeuAlaAlaTrpLysIleSerIleProTyrValAspPhePheGluAspProSerSerGlu 100  
 DB 356 CTTCGCTGCATGGAATATAGCATTCATATGATAGACTTTTGGAGATCCCTCTCTGAA 415  
 QY 101 ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120  
 DB 416 AGCAAGAGAGAAAAGAAAGAAATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 475  
 QY 121 ArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyr 140  
 DB 476 AGAGCAGTGTGGACACAGAGCTGAACATTCCTGTCTATAGAGATATCTTGAATCTCAT 535  
 QY 141 ValLeuGluSerLysLeuThrGluPheHisGlyValAlaPheProAspAlaGlnLeuProSer 160  
 DB 536 GTACTTGAATCAAACTACAGAAATTCATGTGTGATTTCTTGTGATCCAGCTTCTCTCT 595  
 QY 161 LysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerIleArgGluGluPheGln 180  
 DB 596 AAGAGGATCATTCGCCCAAAATATATGAATCTTAAGTCAAAGAGGAGAGAGTCCAA 655  
 QY 181 GluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
 DB 656 GAATATCTACAGAAATCTTCGCAGCATCCAGACTCAGTAATAGTCAACTTCTGCGAGAC 715

201 PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220  
 DB 716 TTTCCTTCCCTTAATGGTGGGAAACACAAATTTCTTGATAGATACTACCAGATCAAT 775  
 QY 221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu 240  
 DB 776 CTGGGAAATATATAAATCTGTCTCTGAAATCAATGAAGAGAGAAAGGTGAGCATTTG 835  
 QY 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
 DB 836 GAACCTTTTATCATGAATTTCAATTAATCTTGTGAGTCTCCAAAGCCTTAACCAAGTAGA 895  
 QY 261 ProGluLeuThrIleLeuSerProThrSerGluAsnLysLysLeuPheAsnAspLeu 280  
 DB 896 CAGAACTGACCAATCTCAGCCCTACTTCCAGAAAACACAAAGAGCTTTCAATGATCTG 955  
 QY 281 PheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGluAsnGlnAsnTyrPhe 300  
 DB 956 TTTTAAATTAATGCAAAACCGTCTGAAATACAGAGAGAAAGCAAAATCAGAAATATTAT 1015  
 QY 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
 DB 1016 ATGGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGATGTAGGACGGGTAGTT 1075  
 QY 321 PheGlnValProAspTrpLeuHisIleLeuLeuMetGlyThrArgIleLeuPheLysAsn 340  
 DB 1076 TTCAGGTTCTGACTGAGCTTTCATCATCTCTTAATGGGAACCTCGAATCTCTTTTAAAC 1135  
 QY 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu 360  
 DB 1136 ACCCTGGAAATGATATCTACTGATTACTTCACTGCTAGTGTAACTAGACAGCTATTTCAG 1195  
 QY 361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380  
 DB 1196 CACCGTTTGGTCTCACTCAACACTCTCAGAGATGCTATATTTCTGTGAAACACTGAA 1255  
 QY 381 ProArgSerLeuGlnAspLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400  
 DB 1256 CCTCGCTCTCTCCAAAGATAAGCAAAAGAGCAAAACAGACTTTTGAAGAAATGATGAT 1315  
 QY 401 TyrIleProAspLeuValLysCysIleGlyGluThrLysTyrGluSerIleArg 420  
 DB 1316 TACATTCAGATCTGTAGTCAAGTGTATTGGTGAAGAAACCAAGATATGAAGCATCAGA 1375  
 QY 421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
 DB 1376 CTTCCTGTTGATGCTTACAGCAACCACTACTCAACAAGCAGCTGACTTATGTTTATTG 1435  
 QY 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
 DB 1436 GACATTTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGGAAGTTACCTCT 1495  
 QY 461 ValThrSerTrpMet 465  
 DB 1496 GTGACATCTTGGATG 1510  
 RESULT 2  
 AY044865 2661 bp mRNA linear PRI 09-SEP-2001  
 LOCUS Homo sapiens sorting nexin 14 (SNX14) mRNA, complete cds.  
 DEFINITION  
 ACCESSION AY044865  
 VERSION AY044865.1 GI:15529063  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2661)  
 AUTHORS Hong W.  
 TITLE The complete coding region of SNX14  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2661)  
 AUTHORS Hong, W.

Direct Submission  
 TITLE Submitted (11-JUL-2001) Membrane Biology Laboratory, Institute of  
 JOURNAL Mol & Cell Biol, 30 Medical Drive, Singapore 117609  
 FEATURES Location/Qualifiers  
 source  
 1. 2661  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 1. 2661  
 /gene="SNX14"  
 1. 2661  
 /note="contains FX domain"  
 /codon\_start=1  
 /product="sorting nexin 14"  
 /protein\_id="AAK9796.1"  
 /db\_xref="GI:15529064"

## gene

## CDS

/translation="MGQKQKRLDVGREICROYPLFCFLLLCLSAASLLNRYIHI  
 LMIFWVAGVVTVCISGPDGLPNIPFTIKYKPKQLQELPQGHSCAVCGKVKC  
 KHRPSLLLENVQWLDIKISKVDASLSEVDIPSIITKLLKAAMKHIEIVKAROK  
 VNTPELQALAEVGPFLHVALSRDELHVLKJELLFPYILPKATDCBSLTL  
 IREILUSGVFLPSLDPLADPTVNHLLIIFIDDPPEKATEPASLVPFVLPQKAEPRN  
 KPSVLEKLEQIREQQLDFRNFLKQEGAVHLOFCLTVEEFNDRIILRPELNDDE  
 MSLHEELQIKYIKYICLDSEIDKIRFDPIVEEIQRIAEAGFYIDVVKLQWTRCLFERY  
 EHLVLENTVTPMFCSDSEYFROLLRGAESPTRNSKLRNTQKRGESFGTSRIGSKI  
 KGVFSTMEGAMLPNYGAEEDDFIEEGIVMEDDSPEVAESTVPENLAWKIS  
 IPYVDFDPSSEKKEKRIEPCVIDVERNDRAVGEPEHNSVYRILEFFVYLESK  
 LTFEGEPDAQLPSKRIIGPNVEFLSKREBEFQYLLQKLPHELSNLSOLLADPUS  
 PNGGTQFLDKLIPDNVIGKI KSVPGKMEKQGHLPKLFIMNFINSCEPFPKPSRP  
 ETLILSPSENNKKLFNLFKNNAENRTERKQNYFMVEVTVEGVYDLYMTVGRV  
 VFQVDFLHLHLLGTRILFKNTLEMYTDYILQCKLEOLFQEHRLVSLITLRLDAIPCE  
 NTEPSLDQKQKQAKTEEMMYIPDLLVVKICIGETKYESIRLLFPGLQOPVLNKLQ  
 ITVLDDIVIQRIEFLPFLNKKVQKEVTSVSNM"

BASE COUNT 855 a 490 c 534 g 782 t

## Alignment Scores:

Pred. No.: 0 Length: 2661  
 Score: 455.00 Matches: 455  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.85% Indels: 0  
 DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x AY044865 (1-2661)

Qy 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLys 30  
 Db 1294 AGAACACACAGAAAGGGAGAAATCATTTGGAAATCAGCAGATAGGTAGCAAAATATA 1353  
 Qy 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50  
 Db 1354 GGGATATTCAAAGTACCACATGAGGAGGAGCTATGTTGCCATTAATTATGGTAGCTGAA 1413  
 Qy 51 GlyLysAspPheIleGluGluGlyLeValValMetGluAspAspSerProValGlu 70  
 Db 1414 GGTGAAGATGATTTTATTGAAGAAGTATTGTTGTAATGGAAGATGATTCCTCCAGTGGAG 1473  
 Qy 71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyr 90  
 Db 1474 GCTGTGAGCACACCTAATACTCCCGAAACCTTGTCTGATGGAATAATAGCATTTCCATAT 1533  
 Qy 91 ValAspPheGluAspProSerGluAspGlyLysGluLysLysGluArgIleProVal 110  
 Db 1534 GTAGATTTTITGAGATCCCTCTCTGAAAGAGGAGGAGAAAGAAAGAAATTCCTGTG 1593  
 Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrp 130  
 Db 1594 TTTTGTATTGATTTTGAAGAATAATGATAGACAGCAGTGGACACAGCCTGAAACATTGG 1653  
 Qy 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
 Db 1654 TCTGTCTATAGAAGATATCTTGAATTTCTATGTACTTGAATCAAACTAACAGAAATTTTCAT 1713

Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
 Db 1714 GGTGCATTTCTTCATGCCAGCTTCCTCTTAAGAGGATCATTTGGCCCCCAAAATATGAA 1773  
 Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisPro 190  
 Db 1774 TTCTTAAAGTCAAAAGAGGAGAGTTCCAAGAAATATCTACAGAAATCTCTGCAGATCCA 1833  
 Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyLysGluThrGln 210  
 Db 1834 GAAGTGAAGTAATGATCACTTCTGGCAGACTTCTTCCCTTAATGGTGGGGAACACAA 1893  
 Qy 211 PheLeuAspLysIleLeuProAspValLeuGlyLysIleIleLysSerValProGly 230  
 Db 1894 TTTCTTGAAGTACTACACAGATGTAAATCTTGGGAAATATAAAATCTCTCTCTGGA 1953  
 Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleLeuSer 250  
 Db 1954 AAATAATGAAAGAGAGAGGTGAGCATTTTGGAAACCTTTTATCATGAATTTCAATATCT 2013  
 Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
 Db 2014 TGTGAGTCTCCAAAGGCTAAACCAAGTAGACAGCAACTGACCATTTCTCAGCCCTACTTCA 2073  
 Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
 Db 2074 GAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATAAGCAACCGTCTGAAAT 2133  
 Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
 Db 2134 ACAGAGAGAAAGCAAAATCAGAAATATTATTTATGGAGTGATGACTGTAGAGGAGTCTAT 2193  
 Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisLeu 330  
 Db 2194 GATTACTGATGTATGTAGGACGGTAGTTTTCAGAGTTCCTGACTGGCTTTCATCATCTC 2253  
 Qy 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
 Db 2254 TTAAATGGGAACCTCGAATCTCTTTTAAACACCCCTGGAAATGTATACTGATTAATCTT 2313  
 Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
 Db 2314 CAGTGTAACTAGAACAGCTATTTTCAGGAGCAGCCGTTTGCTCTCACTCATACACATCTC 2373  
 Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
 Db 2374 AGAGATGCTATATTTCTGTGAAACACTGAACTCGCTCTCTCCAGATTAAGCAAAAGGA 2433  
 Qy 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
 Db 2434 GCAAAACAGACTTTTGAAGAAATGATGAATTAATTCATTCAGATCTGTAGTCAAGTGTAT 2493  
 Qy 411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
 Db 2494 GGTGAAGAAACCAAGATGAAGCATCAGACTTCTGTTTGTGGCTTACAGACACAGTA 2553  
 Qy 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuPheProGlu 450  
 Db 2554 CTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGAACTGTTTCCAGAG 2613  
 Qy 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
 Db 2614 CTCAATAGGTACAAAGGAAGTACTCTGTGACATCTTGGATG 2658

## RESULT 3

BC005110 3038 bp mRNA linear PRI 12-JUL-2001  
 LOCUS Homo sapiens, clone MGC:13217 IMAGE:3959086, mRNA, complete cds.  
 DEFINITION  
 ACCESSION BC005110  
 VERSION BC005110.1 GI:13477272  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)



ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (Bases 1 to 3038)
JOURNAL	Submitted (26-MAR-2001) National Institutes of Health, Mammalian
	Gene Collection (MGC) Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>
	Tissue Procurement: ATCC
	CDNA Library Preparation: Rubin Laboratory
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
	DNA Sequencing by: Genome Sequence Centre,
	BC Cancer Agency, Vancouver, BC, Canada
	<a href="mailto:info@bcsc.bc.ca">info@bcsc.bc.ca</a>
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
	Leticia Hsiao, Martin Krywinski, Reta Kutsche, Oliver Lee, Soo
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
	Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
	Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Scott,
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
	George Yang, Scott Zuyderduyn, Marco Marra.
	Clone distribution: MGC clone distribution information can be found
	through the I.M.A.G.E. Consortium/ILNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>
	Series: IPAL Plate: 18 Row: n Column: 23
	This clone was selected for full length sequencing because it
	passed the following selection criteria: Hexamer frequency ORF
FEATURES	analysis.
source	Location/Qualifiers
	1..3038
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="MGC:13217 IMAGE:3959086"
	/tissue_type="Placenta, choriocarcinoma"
	/clone_lib="NIH MGC 21"
	/lab_host="DH10B-R"
	/note="Vector: pOTB7"
	189..2849
	/codon_start=1
	/product="Unknown (protein for MGC:13217)"
	/protein_id="AAH05110.1"
	/db_xref="GI:13477273"
	/translation="MGQKQKRLRLDVGREICROYPLFCFLLCLSAALLNRYHI
	LMFWSVAGVVFYCSIGPSLLPNFTTIKYKQGLQELFPQGHSCAVCGKVC
	KRRPSLLLENYQWLDKISSKVDASLSEVDIPSIITKLLKAAKHIEVIVKARQK
	VNTEFLQAALEEVGPRLHVALRSRDELHVLKLTLLPPYILPKATDCRSLLTL
	IREILSGSVFLPSLDPLADPTVNHLLIIFDDSPPKATEPASLPVPLQKAPEN
	KXPSVLKELQIRQQLRPFNVLKQEGAVHVLQFLIVEEFNDRIIRLPSELNDE
	MLSEELQKIKYTCDESIDKIFDPFIVEIORAEGYIDVVKLQIRCLFEAY
	EHVLSLENVTFPFCHSDEYFQRLRGASFPTRSKLNRTQKRGSPGISRGSKI
	KGVFSTMBGAMLPNYGVAGEDDFIEGIVMEDDSPVLEAVSTPNTPRNLAAWKIS
	LPYVDFPDPSSEKKEKERIPVTCIDVERNDRAVGEPEHMSVRYLYEFLVLESK
	LTFEGAPFDQALPKSRKILGPKNYEFLSKREBFQYELKLLQHPLSNSQLLADFLS
	PNGGTQFLDKLPDVLNGLIKSVFGKLMKEKQHLFPFLMNFINSCESPKPSRP
	ELTILSPSTNNKKLFNLFKNANRAENTERKQNYFMWEVTVEGYDILMYGVRV
	VQVPLDLMHLMGRLFKNTLEMTYDYLQCKLEQLFQEHRLVLSVLTLLRDAIFCE
	NTEPSLODKOKAQOTKEEMNNTYIPDLLVKICGIBETKYESIRLLFDGLQOPVLNKLQ
	TYVLDIVIQELFPFLNKVQEVTSVSM"
BASE COUNT	949 a 564 c 642 g 883 t
ORIGIN	
Alignment Scores:	
Pred. No.:	0 Length: 3038
Score:	455.00 Matches: 455
Percent Similarity:	100.00% Conservativeness: 0
Best Local Similarity:	100.00% Mismatches: 0

Qy 351 GlnCysLysLeuGlnGlnLeuPheGlnGlnLysArgLeuValSerLeuLeuThrLeuLeu 370  
 Db 2502 CAGTGTAAACTAGAACAGCTATTTCAGGAGCAGCGTTTGGTCTCACTCAACACTTCTC 2561  
 Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
 Db 2562 AGAGATCTATATTCTGTGAACACACAGACCTCGCTCTCTCCAGATAGCAAAAGGA 2621  
 Qy 391 AlaLysGlnThrPheGlnGlnMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
 Db 2622 GCAACACAGACTTTTGAGGAATGATGATTAATCACTCCAGATCTGTAGTCAAGTGATT 2681  
 Qy 411 GlyGlnGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
 Db 2682 GGTGAAGAACCAAGATGATGAAGCATCAGACTTCTGTGATGCTTACAGCAACAGTA 2741  
 Qy 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGlnLeuPheProGlu 450  
 Db 2742 CTCACAGAGAGCTGACTTATGTTTATTTGACATCTGTGATCAGAACTGTTTCAGAG 2801  
 Qy 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
 Db 2802 CTCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 2846

RESULT 4  
 BC046520 2925 bp mRNA linear PRI 21-FEB-2003  
 LOCUS Homo sapiens, Similar to sorting nexin 14, clone IMAGE:5267454,  
 DEFINITION mRNA.

ACCESSION BC046520  
 VERSION BC046520.1 GI:28461364

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission

JOURNAL  
 Submitted (03-FEB-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amgobcm.tmc.edu](mailto:amgobcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IPAK Plate: 91 Row: A Column: 6.  
 Location/Qualifiers

FEATURES

source  
 1. 2925  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5267454"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH MGC 97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

BASE COUNT 948 a 513 c 579 g 885 t

ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 2925  
 Score: 390.00 Matches: 454  
 Percent Similarity: 99.56% Conservat: 0  
 Best Local Similarity: 99.56% Mismatches: 1  
 Query Match: 83.87% Indels: 2  
 DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x BC046520 (1-2925)

Qy 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysLys 30  
 Db 1381 AGGAAACACACAGAAAGGGAGAAATCAATTTGGATATCAGAGAAATAGGTAGCAAAATATAA 1440  
 Qy 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50  
 Db 1441 GGAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCTTAATTAATGTTAGTACGTAA 1500  
 Qy 51 GlyGluAspAspPheIleGluGluGlyValValMetGluAspAspSerProValGlu 70  
 Db 1501 GGTGAGATGATTATTTAGGAAGGTATTGTTGTAATGGAAGATGATTTCTCCAGTGGAG 1560  
 Qy 71 AlaValSerThrProAsnThrProArgAsnLeuAlaIleThrLysIleSerIleProTyr 90  
 Db 1561 GGTGTGAGCACACCTTAATCTCCAGAAACCTTCTGCTGATGAAATATAGCATTTCCATAT 1620  
 Qy 91 ValAspPhePheGluAspProSerSerGluArgLysGlnLysLysGluArgIleProVal 110  
 Db 1621 CTAGACTTTTTTGGAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTTCTGTG 1680  
 Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyr 130  
 Db 1681 TTTTGTATTGATTGTAAGAAAGTAATGAGGAGAGCAGTTGGACAGGACCTTGACATTTGG 1740  
 Qy 131 SerValTyrArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
 Db 1741 TCTGTCTATAGAGATATCTTGAATTCATCTACTTGAATCAAACTCAACAGAAATTTTCAT 1800  
 Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
 Db 1801 GGTGCATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAATTTATGAA 1860  
 Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
 Db 1861 TTTCTTAAAGTCAAAGGAGGAGAGTTCCAGAAATATCTACAGAAACTTCTTCGAGCATCCA 1920  
 Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
 Db 1921 GAACGTAGTATAGTCAACTTCTGCGACACTTTCTTCTCCCTTAATGTTGGGGAACACAA 1980  
 Qy 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGly 230  
 Db 1981 TTTCTTGTATAGATATCTACAGATGTAATCTTTGGGAAATATATAAAATCTGTTCTCTGA 2040  
 Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
 Db 2041 AAATAATGAAGAGAGAGAGTTCAGCATTTGGAACCTTTTATCATGAATTTTCAATTTCT 2100  
 Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
 Db 2101 TGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGACTGACCATTTCTCAGCCCTACTTCA 2160  
 Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
 Db 2161 GAACACACAGAGAGCTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAAAT 2220  
 Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
 Db 2221 ACAGAGAGAGAGAGAGAGTATTTATTTATGAGAGGTGATGACTGTAGAGAGGTCTAT 2280  
 Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHisLeu 330

2281 GATTACCTGATGATAGGACGGTAGTCTTTCCAGGTTCTGACTGGCTTCATCATCTC 2340  
 331 LeuMetGlyThrArgGileuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
 2341 TTAATGGGAACCTCGAATCTCTTTAAACACACCCCTGGAAATGTATCTGATTACTATCTT 2400  
 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuLeu 370  
 2401 CAGTGTAACATAGACACACTATTTCCAGGACACCGTTTGGTCTCCTCATTAACACTTCTC 2460  
 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
 2461 AGAGATGCTATATCTTGTGAACACTGAACCTCGCTCTCCCAAGATAAGCAAAAGGA 2520  
 391 AlaLysGlnThrPheGluGluMetMetAsnTyr-IleProAspLeuLeuValLysCysII 410  
 2521 GCAAAACACAGCTTTGAAGAAATGATGAATG-CATTCCAGATCTGTGTAGTCAAGTGTAT 2579  
 410 eGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVa 430  
 2580 TGGTGAAGAAACCAAGTAGTGAAGCATCAGACTTCTGTGATGGCTTTACAGCAACAGT 2639  
 430 IleuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProG 450  
 2640 ACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGTGTATAGGACTGTTTCCAGA 2699  
 450 uLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
 2700 GCTCAATAGGTACAAAGAGGATTTACCTGTGTGACATCTTGGATG 2745

RESULT 5  
 AF121863 1593 bp mRNA linear PRI 14-SEP-2001  
 LOCUS Homo sapiens sorting nexin 14 (SNX14) mRNA, partial cds.  
 DEFINITION  
 ACCESSION AF121863  
 VERSION AF121863.1 GI:4689265

KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS Teasdale, R.D., Locci, D., Houghton, P., Karlsson, L. and Gleeson, P.A.  
 TITLE A large family of endosome-localized proteins related to sorting nexin 1

JOURNAL Biochem. J. 358 (Pt 1), 7-16 (2001)  
 MEDLINE 21378165  
 PUBMED 11485546  
 REFERENCE 2 (bases 1 to 1593)  
 AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.  
 TITLE Identification of eleven novel human sorting nexin molecules. A sub-group of the sorting nexin family is associated with the early endosomes

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1593)  
 AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-JAN-1999) The R.W. Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, CA 92121, USA

FEATURES  
 source  
 1..1593  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 <1..1593  
 /gene="SNX14"  
 <1..1125  
 /gene="SNX14"  
 /codon\_start=1  
 /product="sorting nexin 14"  
 /protein\_id="AA027836.1"  
 /db\_xref="GI:4689266"

gene  
 CDS

/translation="DFFDPSSERKEKRIIPFCIDVNRDRAVGEPEHWSVYRR  
 YLEFVLSEKITEFHGAPPDAQLPSKRIIGPKNYEFLKSEEEFQYLOKLQHPELS  
 NSQLADFLSPNGGETOFLDKILPDVNLKIKVPGKLMKEKQHLPEFPTMNFINS  
 ESPKPRPELTILSPSENKKLNDLTKNNARAEATERKONQVEMEVMTVEGV  
 IDILMYGRVVFQFDWLHLHMGTRILFKNTLEMYTDYLOCKLEQLFQHRVLSLI  
 LDALDAICENTPEPSQDKQKGAQFPEEMNYIPDLLVKICIGETIKYISIRLLFDG  
 LQQPVINKQLTYVLLDIVIQLPELNKVKQEVTSVTSWM"

BASE COUNT 533 a 256 c 294 g 510 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 1593  
 Score: 374.00 Matches: 374  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 80.43% Indels: 0  
 DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x AF121863 (1-1593)

QY 92 AspPheGluAspProSerSerGluArgLysGluLysGluArgIleProValPhe 111  
 DB 1 GACTTTTTTGGAGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGTTT 60  
 QY 112 CysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSer 131  
 DB 61 TGTATTGATGTTGAAGAAATGATAGAGAGAGAGTTGGACAGAGCTGAACATTGGTCT 120  
 QY 132 ValTyrArgArgTyrIleuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGly 151  
 DB 121 GTCTATAGAAGATATCTTGAATCTATGTACTTGAATCAAAACTAACAGAAATTCATGTT 180  
 QY 152 AlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPhe 171  
 DB 181 GCATTTTCTGATGCCAGCTTCTCTTAAAGAGAGATTCATGGCCCCCAAAATATGAATTC 240  
 QY 172 LeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisProGlu 191  
 DB 241 TTAAGTCAAGAGGAGAGAGTTCCAGAAATATCTACAGAAACTTCTGCAGCATCCAGAA 300  
 QY 192 LeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPhe 211  
 DB 301 CTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGGTGGGAAACACAAATTT 360  
 QY 212 LeuAspLysIleLeuProAspValLeuGlnLysIleIleLysSerValProGlyLys 231  
 DB 361 CTTGATAAGATACCTACAGATGTAAATCTTGGGAAATATATAAATCTCTTCTCTGAAA 420  
 QY 232 LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCys 251  
 DB 421 CTAAATGAAGAGAGAGAGTCAAGCATTTGGAACTTTTATCATGAATTTCAATTTCTGT 480  
 QY 252 GluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGlu 271  
 DB 481 GAGTCTCCAAAGCCTAAACCAAGTAGACACAGAACTGACCATTCTCAGCCTACTTCAGAA 540  
 QY 272 AsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThr 291  
 DB 541 AACCAACAGAGAGCTTTTCAATGATCTGTTTAAAAATAATGCAAAACCGTCTCGAAAA 600  
 QY 292 GluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAsp 311  
 DB 601 GAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAGAGAGTCTATGAT 660  
 QY 312 TyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisLeuLeu 331  
 DB 661 TACTGATGATGTAGAGCGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTA 720  
 QY 332 MetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGln 351  
 DB 721 ATGGGAACTCGAATCTCTTTTAAAAACACCCCTGGAATGTATATCTGATTACTATCTTCAG 780  
 QY 352 CysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArg 371



```
Db 2917 GATTACCTGATGATGATAGGACGGGTAGTTTCCAGGTTCTCGACTGGCTTCATCATCTC 2976
Qy 331 LeuMetGlyThrArgIleLeuPheLeuAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350
Db 2977 TTAATGGGAAGTCAATCCCTTTTAAAAACACCCCTGGAATGATATCTACTACTATCTT 3036
Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370
Db 3037 CAGTGTAACCTAGAACAGCTATTTCAGGACACCGTTTGTCTCTCACTCATACACTTCTC 3096
Qy 371 ArgAspAlaIlePheCysGluAsnThr-GluProArgSerLeuGlnAspLysGlnLysG 390
Db 3097 AGAGATGCTATCTGTGAAACCC-TGAACCTCGCTCTCTCCAAAGTAAGCAAAAGG 3155
Qy 390 ValIleGlnThrPheGlnGluMetMetAsnTyrIleProAspLeuValLysCysIle 410
Db 3156 AGCAAAACAGACTTTTGAAGAAATGATGAATTACATTCAGATCTGTGTAGTCAAGTGAT 3215
Qy 410 eGlyGluGluThr 414
Db 3216 TGGTGAGAAACC 3228

RESULT 7
LOCUS AX512835 3145 bp DNA linear PAT 03-OCT-2002
DEFINITION Sequence 4 from Patent WO02062839.
ACCESSION AX512835
VERSION AX512835.1 GI:23504019
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Daemen, M.J., Cleutjens, C.B. and Zaman, G.J.
Markers of unstable atherosclerotic plaques
Patent: WO 02062839-A 4 15-AUG-2002;
Universiteit Maastricht (NL)
Location/Qualifiers
1. 3145
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 985 a 570 c 659 g 931 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 3145
Score: 330.00 Matches: 430
Percent Similarity: 99.54% Conservative: 0
Best Local Similarity: 99.54% Mismatches: 1
Query Match: 70.97% Indels: 2
DB: 6 Gaps: 0

US-09-744-313A-1 (1-465) x AX512835 (1-3145)
Qy 35 SerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspAsp 54
Db 1673 AGTACACACATGGAGGAGCTATGCTTAATTATGTTGTAGCTGAAGGTGAAGATGAT 1732
Qy 55 PheIleGluGluGlyIleValValMetGluAspSerProValGluAlaValSerThr 74
Db 1733 TTTATTGAAGAGGTATTGTTGTAATGAAGATGATTCTCCAGTGGAGGCTGTGAGCACA 1792
Qy 75 ProAsnThrProArgAsnLeuAlaAlaTyrLysIleSerIleProTyrValAspPhePhe 94
Db 1793 CCTAATATCTCCCGAAACCTTCTCGATGGAAATATGATTCATTCATATGATAGATTTTT 1852
Qy 95 GluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAsp 114
Db 1853 GAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1912
Qy 115 ValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArg 134
```

RESULT 8  
AK000362

```
Db 1913 GTTGAAGAAATGATAGAGAGCAGTTGGACACAGCCTGAACATTTGGTCTGCTATAGA 1972
Qy 135 ArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPhePro 154
Db 1973 AGATATCTTGATTTCTATGTTACTTGAATCAAACTACAGAAATTTTCATGGTCAATTCCT 2032
Qy 155 AspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSer 174
Db 2033 GATGCCACAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAATATGAAATTTCTTAAGTCA 2092
Qy 175 LysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsn 194
Db 2093 AAGAGGGAAGAGTTTCCAAGATATCTACAGAAATCTCTGAGCAGATCCAGAACTGAGTAAT 2152
Qy 195 SerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLys 214
Db 2153 AGTCAACTCTTGCCAGACTTTCTTTCCCTTAATGGTGGGAAACACAAATTTCTTGATAAG 2212
Qy 215 IleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLys 234
Db 2213 ATACTACCAGATGTAAATCTTTGGGAAATATATAAATCTGTTCCTGGAAAACTAATGAAA 2272
Qy 235 GluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerPro 254
Db 2273 GAGAAAGGTCAGCATTTTGGAACTTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCA 2332
Qy 255 LysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLys 274
Db 2333 AAGCCTAAACCAAGTAGACCAGAACTGACCATTCTGAGCCCTACTTTCAGAAAACACACAG 2392
Qy 275 LysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLys 294
Db 2393 AAGCTTTTCAATGATCTCTTTAAAAATAATGCAACCGTCTGAAATATACAGAGAGAAAG 2452
Qy 295 GlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMet 314
Db 2453 CAAATCAAGATTAATTTATGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGATG 2512
Qy 315 TyrValGlyArgValValPheGlnVal-ProAspTrpIleuHisIleLeuLeuMetGlyTh 334
Db 2513 TATGTAGGACGGGTAGTTTTCAGAT-TCCTGACTGGCTTCATCATCTCTTAATGGGAAC 2571
Qy 334 rArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLe 354
Db 2572 TCGAATCTCTTTTAAAAACACCCCTGGAATGTATACTGATTAATCTCTCAGTGTAACT 2631
Qy 354 uGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaI 374
Db 2632 AGACAGCTATTTTCAGGAGCACCGTTTGTCTCTCACTCATAACACTTCTCAGAGATGCTAT 2691
Qy 374 ePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnTh 394
Db 2692 ATTCTGTGAAAACACTGAAACCTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACGAC 2751
Qy 394 rPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluTh 414
Db 2752 TTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAAC 2811
Qy 414 rLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysLe 434
Db 2812 CAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCCAGTACTCAACAAGCA 2871
Qy 434 nLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysVa 454
Db 2872 GCTGACTTATGTTTATTTGACATTTGTGATACAGAACTGTTTCCAGAGCTCAATAAGT 2931
Qy 454 lGlnLysGluValThrSerValThrSerTrpMet 465
Db 2932 ACAAAGGAAGTTACCTCTGTGACATCTTGGATG 2965
```

LOCUS AK000362 3145 bp mRNA linear PRI 22-FEB-2000  
 DEFINITION Homo sapiens cDNA FLJ20355 fis, clone HEP15804, highly similar to  
 AF121863 Homo sapiens sorting nexin 14.  
 ACCESSION AK000362  
 VERSION AK000362.1 GI:7020397  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,  
 Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,  
 Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
 Nakamura, Y., Isogai, T. and Sugano, S.  
 TITLE NEDO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3145)  
 AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
 Shibahara, T., Tanaka, T. and Nakamura, Y.  
 DIRECT SUBMISSION  
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,  
 Minato-ku, Tokyo 108-8639, Japan [E-mail:cdna@ims.u-tokyo.ac.jp,  
 Tel:81-3-5449-5286, Fax:81-3-5449-5416]  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing; Department of  
 Virology and Human Genome Center, Institute of Medical Science,  
 University of Tokyo (partly supported by Science and Technology  
 Agency).  
 FEATURES  
 source Location/Qualifiers  
 1..3145  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="HEP15804"  
 /cell\_line="HepG2"  
 /cell\_type="hepatoma"  
 /clone\_lib="HEP"  
 /notes="cloning vector pBE18SFL3"  
 misc\_feature 1..3145  
 /notes="highly similar to AF121863 Homo sapiens sorting  
 nexin 14"  
 BASE COUNT 985 a 570 c 659 g 931 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 3145  
 Score: 330.00 Matches: 430  
 Percent Similarity: 99.54% Conservative: 0  
 Best Local Similarity: 99.54% Mismatches: 1  
 Query Match: 70.97% Indels: 2  
 DB: 9 Gaps: 0  
 US-09-744-313A-1 (1-465) x AK000362 (1-3145)  
 QY 35 SerThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspAsp 54  
 DB 1673 AGTACCACATGAGGAGGAGCTATGTTGCCCTAAATTATGGTGTAGCTGAAGGTGAAGATGAT 1732  
 QY 55 PheIleGluGlyIleValValMetGluAspSerProValGluAlaValSerThr 74  
 DB 1733 TTATTGAAGAAGTATTGTTTAATGAAGATGATTCCTCAGTGGAGGCTGTGAGCACA 1792  
 QY 75 ProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPhePhe 94  
 DB 1793 CCTAATACTCCCGAAACCTTGTCATGGCAAAATTAGCATTCATATGATGACTTTT 1852  
 QY 95 GluAspProSerGluArgLysGlyLysLysLysLysLysLysLysLysLysLys 114  
 DB 1853 GAGGATCCCTCTGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1912

QY 115 ValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArg 134  
 DB 1913 GTTGAAGAAATCATAGAAGAGCAGTTGGACAGCAGCCTGCAACATTTGGTCTCTCTATAGA 1972  
 QY 135 ArgTyrIleGluPheTyrValIleGluSerLysLeuThrGluPheHisGlyAlaPhePro 154  
 DB 1973 AGATATCTTGAATCTTATGTACTTGAATCAAACTAACAGAATTTTCATGGTGCATTTCT 2032  
 QY 155 AspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSer 174  
 DB 2033 GATGCCAGCTTCTCTTAAAGAGATCATTTGGCCCCAAAATATGAATCTTAAAGTCA 2092  
 QY 175 LysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsn 194  
 DB 2093 AAGAGGGAAGAGTTCCAGAATATCTACAGAACTTCTGCAGCATCCAGAACTGAGTAAT 2152  
 QY 195 SerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLys 214  
 DB 2153 AGTCAACTCTGTCAGACTTTCTTCCCTTAATGGTGGGAAACACAAATTTCTTGTATAAG 2212  
 QY 215 IleLeuProAspValAlaLeuGlyLysIleLysSerValProGlyLysLeuMetLys 234  
 DB 2213 ATACTACCAGATGTAATCTTGGGAAATATATAAAATCTGTTCTCGAAACTAATAGAA 2272  
 QY 235 GluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerPro 254  
 DB 2273 GAGAAAGTCAGATTTGGACCTTTATCATGAATTTCAATATTTCTTGTGAGTCTCCA 2332  
 QY 255 LysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnLys 274  
 DB 2333 AAGCCTAAACCAAGTAGACCAGAACTGACCATTCTCAGCCCTACTTTCAGAAACAACAAG 2392  
 QY 275 LysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLys 294  
 DB 2393 AAGCTTTTCAATGATCTGTTTAAAAATAATGCAAAACCGTCTGAAATATACAGAGAAAG 2452  
 QY 295 GluAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMet 314  
 DB 2453 CAAAATCAGAAATTTATTTATGGAGGTGATGACCTGAGAGAGTCTATGATTTACCTGATG 2512  
 QY 315 TyrValGlyArgValValPheGlnVal - ProAspTyrLeuHisIleLeuMetGlyLys 334  
 DB 2513 TATGTAGACGGGTAGTTTTCAGAT - TCTGTACTGGCTTCATCACTCTTATGGGAAC 2571  
 QY 334 rArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrIleGlnCysLysLe 354  
 DB 2572 TCGAATCTCTTTAAAAACACCCCTGGAAATGTATATGATTACTATCTTCTCAGTGTAAC 2631  
 QY 354 uGluGlnPheGlnGlnHisArgLeuValSerLeuIleThrIleLeuArgAspAlaI 374  
 DB 2632 AGAACAGCTATTTTCAGAGCAGCCGTTGGTCTCACTCATTAACACTTCTCAGAGATGCTAT 2691  
 QY 374 ePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnTh 394  
 DB 2692 ATTCTGTGAAACACTGAACCTGCTCTCTCCAGATTAACAAAAGGAGCAACAAACAGAC 2751  
 QY 394 rPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluTh 414  
 DB 2752 TTTTGAAGAATGATGATTAATACATTCAGATCTGTTAGTCAAGTATTGGTGGAGAAAC 2811  
 QY 414 rLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysG 434  
 DB 2812 CAAGTATGAAGCATCAGACTTCTGTTTGTGGTGTACAGCAACCCAGTACTCTCAACAAGCA 2871  
 QY 434 nLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysVa 454  
 DB 2872 GCTCACTATGTTTATTTATGGACATTTGATACAGAACTGTTTCCAGAGCTCAATAAGGT 2931  
 QY 454 lGlnLysGluValThrSerValThrSerTrpMet 465  
 DB 2932 ACAAAGGAGTACCTCTGTGACATCTTGGATG 2965

RESULT 9	AK095380	3616 bp	mRNA	linear	PRI 15-JUL-2002
LOCUS	AK095380				
DEFINITION	Homo sapiens cDNA FLJ38061 fis, clone CTONG2014966, highly similar to SORTING NEXIN 14.				
ACCESSION	AK095380				
VERSION	AK095380.1	GI:21754626			
KEYWORDS	oligo capping, fis (full insert sequence).				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuoka, M., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Nagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Ohshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isoigai, T.				
TITLE	NEDO human cDNA sequencing project				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3616)				
AUTHORS	Isoigai, T. and Yamamoto, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-JUL-2002) Takao Isoigai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan				
COMMENT	(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.				
FEATURES	Location/Qualifiers				
source	1. 3616				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CTONG2014966"				
	/tissue type="tongue, tumor tissue"				
	/clone lib="CTONG2"				
	/note="cloning vector: pME18SPL3"				
BASE COUNT	1078 a	629 c	710 g	1199 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	3,28e-291	Length:	3616		
Score:	269.00	Matches:	453		
Percent Similarity:	98.91%	Conservative:	0		
Best local Similarity:	98.91%	Mismatches:	2		
Query Match:	57.85%	Indels:	5		
DB:	9	Gaps:	0		
US-09-744-313A-1 (1-465) x AK095380 (1-3616)					
QY	11	ArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgIleGlySerLysIleLys	30		
Db	2080	AGCAACACACAGAAAAGGGAGATCAATTCGAAATCAGCAGATAGGTAGCAAAATTA	2139		
QY	31	GlyValPheLysSerThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu	50		
Db	2140	GGAGTATTCAAAAGTACCAATGAGGAGGAGCTATGTTGCCCTAAATATGGTGTAGCTG	2199		
QY	51	GlyGluAspPheIleGluGlyIleValValMetGluAspSerProValGlu	70		
Db	2200	GGTGAGATGATTTATTGAAGAAGTATTGTTGTAATGGAGAGATTCCTCCAGTGGAG	2259		
QY	71	AlaValSerThrProAsnThrProArgAsnLeuAlaTrpLysIleSerIleProTyr	90		

Db	2360	GCTGTGACACACCTAATACTCCCGAAACCTTGTGTCATGGAAATTAGCATTCATAT	2319		
QY	91	ValaspPheGluAaspProSerSerGluArgLysGlu-LysLysGluArgIleProVa	110		
Db	2320	GTAGACTTTTGTAGGATCCCTCTCTGAAGAGAGAGAGAGAGAGAGATCTCTGT	2379		
QY	110	lPheCysIleAaspValGluArgAsnAaspArgAlaValGlyHisGluProGluHisI	130		
Db	2380	CTTTTGTATTGATGTTGAAGAAGATGATAGAGAGAGCTTTGGACACGAGCTGAAC	2439		
QY	130	pSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	150		
Db	2440	CTCTGTCTATAGAAGATATCTTGAATCTATGATCTTGAATCAAACTAAACAGAA	2499		
QY	150	sGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr	170		
Db	2500	TGGTGCACTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGCCCCCAAAATTA	2559		
QY	170	uPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGln	190		
Db	2560	ATTCTTAAGTCAAGAGGAGAGAGTCCAGAGATATCTACAGAACTTCTGCGAG	2619		
QY	190	oGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGlu	210		
Db	2620	AGAAGCTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTAATGTTGGGAA	2679		
QY	210	nPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro	230		
Db	2680	ATTTCTTATAGTAACTTACCAGATGTAAATCTTTGGGAAATTAATAAATCTG	2739		
QY	230	YValLeuMetLysGluLysGlyGlnHisLeuGluPro-PheIleMetAsnPheIle	250		
Db	2740	AAATCTAATGAAGAAGAGAGGTGAGCATTTGGAAT-TTTTATCATGAATTTCA	2798		
QY	250	erCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerPro	270		
Db	2799	CTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCACTTCTG	2858		
QY	270	erGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArg	290		
Db	2859	CAGAAACACAAAGAGAGCTTTTCAATGATCTGTTTAAATAATAATGCAACCG	2918		
QY	290	snThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGlu	310		
Db	2919	ATACAGAGAGAGAGAGAGATCAAGAAATTTATTCGAGGTGATGACTGTAG	2978		
QY	310	YzAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeu	330		
Db	2979	ATGATTTACTGATGATGTATGATAGGACGGTAGTTTCCAGGTTCTGACTG	3038		
QY	330	euLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAsp	350		
Db	3039	TCTTAATGGGAACTCGAATCTCTTTTAAACACACCGCTGAAATGTATACT	3098		
QY	350	euGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThr	370		
Db	3099	TTGAGTGTAAATAGAACAGCTATTTCAGGAGACCGGTTGGTCTCCTCAT	3158		
QY	370	euArgAspAlaIlePheCysGluAsnThrGlu-ProArgSerLeuGlnAspLys	389		
Db	3159	TCAGAGATGCTATATTCTGTGAAACACTCA-ACCTCGCTCTCTCAAGATA	3217		
QY	390	GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuValLys	409		
Db	3218	GGAGCAAAACACAGCTTTTGAAGAATGATGAATTACATTCAGATCTGTAG	3277		
QY	410	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGln	429		
Db	3278	ATTGGTGAAGAAACCAAGATGAAAGCATCAGACTTCTGTTTGTATGGCTT	3337		
QY	430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuPhe	449		
Db	3338	GTAAGCAAGAGCTGACTTATGTTTATTGGACATTTGTGATACAGAACTG	3397		



QY	403	ProAspLeuLeuValIysCysIleGlyGluGluThrIysTyCgluSerIleArgLeuLeu	422
Db	321	CCAGATCTGTAGTCAAGTGTATGTGTGAAGAACCAAGTATGAAGCATCAGACTTCTG	380
QY	423	PheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyValleuLeuAspIle	442
Db	381	TTTGATGGCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTTTTATTGGACATT	440
QY	443	ValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSerValThr	462
Db	441	GTGATACAGGAAGTCTGTTTCCAGAGCTCAATAAGGTACAAAGGAAGTTACTCTCTGTGACA	500
QY	463	SerTyrMet 465	
Db	501	TCTTGGATG 509	
RESULT 11			
BD113855			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
OS Homo sapiens (human)			
PN JP 2002010789-A/5932			
PD 15-JAN-2002			
PF 07-AUG-2000 JP 2000280989			
PR 05-AUG-1999 US 60/147499			
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI			
GIORDANO			
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC			
C12N1/21,			
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC			
C12N15/00			
CC EST and encoded human protein			
FH Key			
FT source			
FT Location/Qualifiers			
1. 425			
/organism='Homo sapiens (human)'			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.: 2,65e-148 Length: 425			
Score: 141.00 Matches: 141			
Percent Similarity: 100.00% Conservative: 0			
Best Local Similarity: 100.00% Mismatches: 0			
Query Match: 30.32% Indels: 0			
DB: 6 Gaps: 0			
US-09-744-313A-1 (1-465) x BD113855 (1-425)			
QY	297	GlnAsnTyrPheMetGluValMetThrValIleGlyValTyrAspTyrLeuMetTyrVal	316
Db	1	CAGAAATTATTTATGAGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGATGATGTA	60
QY	317	GlyArgValValPheGlnValProAspTyrPheHisHisLeuLeuMetGlyThrArgIle	336
Db	61	GGACGGGTAGTTTTCCAGGTTCTCTGATGGCTTCATCATCTCTTAATGGGAACTCGAATC	120

337 LeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGln 356  
121 CTCTTTAAACACACCTGGAAATGATATACCTGATTTACTATCTTCAGTCTAACTAGAACAG 180  
357 LeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaIlePheCys 376  
181 CTATTTTCAGAGACACCGTTTGGTCTCACTCATACACTCTCTCAGAGATGCTATATTCTGT 240  
377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyValLysGlnThrPheGlu 396  
241 GAAACACACTGAACCTCGCTCTCTCAAGATAAGCAAAAGAGCAAAACAGACTTTTGAA 300  
397 GluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr 416  
301 GAATGATGATTACATTCAGACTCTGTAGTCAAGTGTATTGGTGAAGAACCAAGTAT 360  
417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436  
361 GAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACT 420  
437 Tyr 437  
421 TAT 423

RESULT 12  
BD026500  
LOCUS  
DEFINITION  
Sequence tag and encoded human protein.  
ACCESSION  
BD026500  
VERSION  
BD026500.1 GI:22567723  
KEYWORDS  
JP 2001269182-A/2746.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 358)  
AUTHORS  
Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.  
TITLE  
Sequence tag and encoded human protein  
JOURNAL  
Patent: JP 2001269182-A 2746 02-OCT-2001;  
GENSET  
COMMENT  
OS Homo sapiens (human)  
PN JP 2001269182-A/2746  
PD 02-OCT-2001  
PF 24-FEB-2000 JP 2000118773  
PR 26-FEB-1999 US 60/122487  
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES  
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10,  
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC  
G06F15/40  
CC  
PH KEY Location/Qualifiers  
FT CDS  
1..358 Location/Qualifiers  
source  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 125 a 63 c 59 g 101 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.16e-123 Length: 358  
Score: 119.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 25.59% Indels: 0  
DB: 6 Gaps: 0  
US-09-744-313a-1 (1-465) x BD026500 (1-358)

216 LeuProAspValAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGlu 235  
2 CTACCAGATGTAATCTCTGGGAAATATATAAATCTGTCTCTGGAAACATAATGAAAGAG 61  
236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLys 255  
62 AAAGTCAGCATTTGGAAACCTTTTATCATGATTTTCAATTAATCTTGTGAGTCTCCAAAG 121  
256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLys 275  
122 CCTAAACCAAGTAGACCACTGACCATTTCTAGCCCTACTCTCAGAAAAACCAAGAAG 181  
276 LeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
182 CTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATATACAGAGAGAAGCAA 241  
296 ArgGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315  
242 AATCAGATTAATTTTATGGAGGTGATGACTAGAGAGGTCTATGATTACCTGATGAT 301  
316 ValGlyArgValAlaPheGlnValProAspTyrLeuHisLeuLeuMetGlyThr 334  
302 GTAGGACGGTAGTTTTCAGGTTCTGACGTGCTTCATCATCTCTTAATGGGAACC 358

## RESULT 13

BC043328  
LOCUS  
DEFINITION  
Mus musculus RIKEN cDNA C330035N22 gene, mRNA (cdna clone MGC:49424  
IMAGE:540015), complete cds.  
ACCESSION  
BC043328  
VERSION  
BC043328.1 GI:27694048  
KEYWORDS  
MGC.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1782)  
AUTHORS  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carrinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A.C., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

## TITLE

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1782)  
Strausberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Heiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Nesse, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Teai, Nasajia van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: iBAK Plate: 86 Row: p Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEATURES

source Location/Qualifiers

1. 1782

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="MGC:49424 IMAGR:5400157"

/tissue type="Eye, retina, mouse strain C57Bl/6"

/clone\_lib="NIH MGC\_94"

/lab host="DH10B"

/note="Vector: PCMV-SPORT6"

gene

1. 1782

/gene="C330035N22Rik"

/note="synonym: B830022K16"

/db\_xref="LocusID:244962"

/db\_xref="MGI:2443156"

228..1616

CDS

/codon\_start=1

/product="C330035N22Rik protein"

/protein\_id="AAH43328.1"

/db\_xref="GI:27694049"

/db\_xref="LocusID:244962"

/translation="MRICLIFKSTQKGSFGIRIGKIKGVKSTMEGAVLPNG  
VAEGDDTIEGIVMBDSFVEASTPTNLAAMKISIPYDFFDPSSEKEXX  
ERIPFVIDVERNDRAVGHPEHWSVRYRLEFVLESLKTEBPHGTFPPAQLPSKRI  
IQKNVYPLKREBFQYLOKLVHPELSQLADPLSPNGGETOPLDKILPDVNL  
GKIIKSPQKMKEGHLEPFINSFINSCKSPKSPBELTILSPSENKLEND  
LFKNANRAENTKQNYFNEVMTVDGVDYLMYGRVVFQPDMLHLLMGTRIL  
FKNTLEMTDYLOCKJQLFOEHLVSLITLLRDAIFCENTBPSLQDKQKAKQTF  
EMMYNIPDLIVKCIQEETKYESIRLLFDGLQQLVINKQLTYLLDVIQELFPELNL  
VQKEATSTMSWN"

BASE COUNT 567 a 339 c 380 g 496 t

ORIGIN

Alignment Scores:

Pred. No.: 8,26e-112 Length: 1782  
Score: 109.00 Matches: 439  
Percent Similarity: 97.34% Conservative: 0  
Best Local Similarity: 97.34% Mismatches: 6  
Query Match: 12 Indels: 12  
DB: 20 Gaps: 0

US-09-744-313a-1 (1-465) x BC043328 (1-1782)

QY 13 ThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLysGlyVal 32  
DB 255 ACACAAAAAGAGGAGAGATCGTTTGGAAATCAGCAGAAATAGGTAGCAAAAATTAAGGCGTA 314  
QY 33 PheLysSerThrThrMetGluGlyAlaMet-LeuProAntTyrGlyValAlaGluGlyG1 52  
DB 315 TTCAAGAGTACGCAATGAGGAGAGCTGT-GCTGGCTTAATTACGGGGGTGGCTGAAGGCGA 373  
QY 52 uAspAspPheIleGluGluGlyIleValValMetGluAspAspSerProValGluAlaVa 72  
DB 374 AGATGACTTTATTTGAAGAGGAGGATTCGTGTATGAGGAGATGACTCTCCAGTAGAAGCTGT 433

QY 72 lSerThrProAntThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValas 92  
DB 434 GAGCACACCTAACTCTCTCGAAACCTTCTGCTCGAAATATTAGCAITTCGATGTAGA 493  
QY 92 pPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCy 112  
DB 494 CTTTTTTGAAGATCCCTCTCTCTGAAAGGAAAGAGAGAGAGAGAAATCTCTGTTTGG 553  
QY 112 sIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerVa 132  
DB 554 TATCGATGTTGAAAGAAACGACAGAGCAGTCTGGACACGAGCCTGAGCATTTGCTGT 613  
QY 132 lTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAl 152  
DB 614 CTATAGAGATATCTTGAATCTTATGTTCTTGAATCAAACTAACAGAAATTCATGGCAC 673  
QY 152 a-PheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTrpGluPheL 172  
DB 674 -ATTCTCTGATGCTCAGCTTCCATCAAAAGATCATTTGCCCCCAAAACTATGAGTTT 732  
QY 172 euLysSerLysArgGluGluPheGlnGluTyrLeuGlnLys-LeuLeuGlnHisProGlu 191  
DB 733 TGAAGTCGAAGAGAGAGAGAGATTCAGGAGTATCTGCAGAAACTTTGTG-CAGCACCGGAG 791  
QY 192 LeuSerAsnSerGlnLeuLeuAlaPhePheLeuSerProAsnGlyGlyGluThrGlnPhe 211  
DB 792 CTGAGTAACAGCCAGCTCTCTGCTGCTACTTCTCTCCCAACGGTGGGAGACACAGATT 851  
QY 212 LeuAspLysIleLeuProAspValLeuGluLysIleIleLysSerValProGlyLys 231  
DB 852 CTAGATAAGATATCTCCAGATGTAATCTTGGAAATAATTAAGTCTCTTCTCTGGAAA 911  
QY 232 LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsn-PheIleAsnSerCy 251  
DB 912 CTAATGAAGAGAAAGAGTCAAGATTTGGAGCCATTCATCATGAG-TTTCATTAATCTTG 970  
QY 251 sGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGl 271  
DB 971 CGAATCTCCAAACCTTAACCGAGTCGCCAGAACTGACGATCTTCAGCCCCACTTCGGA 1030  
QY 271 uAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnTh 291  
DB 1031 AAACAACAAGAGCTTTTCAACGATCTGTTTAAAGATAATGCAAAACCGGCTGAGAACAC 1090  
QY 291 rGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGlu-GlyValTyrA 311  
DB 1091 AGAGCGAAAGCAAAATCAGAACTACTTTCATGGAGGTGATGCTGTGA-CGGAGTCTATG 1149  
QY 311 spTyrLeuMetTyrValValGlyArgValValPheGlnValProAspTrpLeuHisHisLeuL 331  
DB 1150 ATTACCTGATGATGTAGGGCGAGTAGTTTTCAGTCCAGACTCCGCTTCTCATCTTC 1209  
QY 331 euMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuG 351  
DB 1210 TAATGGGAACCTCGAATCCTCTTTAAGAACACACCTCGAAATGTACACAGACTACTACCTCC 1269  
QY 351 lncLysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuA 371  
DB 1270 AGTCGAAGCTGGAGCAGCTGTTTTCAGAGCAGCCGCTGCTCTCCCTCATCCCTTCTCA 1329  
QY 371 rGAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyA 391  
DB 1330 GAGATGCTATATTTGTTGAAAAATCTGAACTCGCTCTCTCCAGATATAGCAAAAAGGAG 1389  
QY 391 lalysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu-VallysCysIle 410  
DB 1390 CAAAAAGAGCTTTTGAAGAAATGATGATATACATTCAGATCTGAT-AGTCAAGTGTATC 1448  
QY 411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
DB 1449 GGTGAAGAAACCAAGTACGAAAGCATCAGACTCTCTGTTGATGGCTGCGAGCAGCCAGTG 1508  
QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450

Db 1509 CTCAACAGCAGCTGACCTATGTTCTGACATTGTGATACAGAGCTGTTTCCAGAG 1568

Qy 451 LeuAenLysValGlnLysGlu 457

Db 1569 CTAATAAGGTACAAAGGAA 1589

RESULT 14

AL589666/c AL589666 129010 bp DNA linear PRI 12-MAY-2001

LOCUS Human DNA sequence from clone RP11-321N4 on chromosome 6, complete

DEFINITION

ACCESSION AL589666

VERSION AL589666.5 GI:14041764

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 15, 2001 this sequence version replaced gi:13751565.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

RP11-321N4 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-321N4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-3017 is at 128911 in this sequence. The true right end of clone RP11-3016 is at 100 in this sequence.

Location/Qualifiers

1..129010

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="RP11-321N4"

/clone\_lib="RPCI-11.2"

2..192

/note="MER21B repeat: matches 3. .207 of consensus"

240..909

/note="MER21B repeat: matches 30. .715 of consensus"

1593..1682

/note="MIR repeat: matches 35. .145 of consensus"

2030..2102

/note="L2 repeat: matches 2681. .2750 of consensus"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region 3459..3574

repeat\_region /note="FLAM C repeat: matches 1. .117 of consensus"

repeat\_region 5635..5670

repeat\_region /note="L1ME repeat: matches 683. .718 of consensus"

repeat\_region 6887..6982

repeat\_region /note="L2 repeat: matches 2616. .2704 of consensus"

repeat\_region 8284..8356

repeat\_region /note="L2 repeat: matches 2639. .2710 of consensus"

repeat\_region 8380..8405

repeat\_region /note="L3 copies 2 mer tt 100% conserved"

repeat\_region 8440..8601

repeat\_region /note="MIR repeat: matches 1. .169 of consensus"

repeat\_region 8825..9360

repeat\_region /note="L2 repeat: matches 1686. .2301 of consensus"

repeat\_region 9423..9592

repeat\_region /note="L2 repeat: matches 2560. .2727 of consensus"

repeat\_region 12088..12175

repeat\_region /note="44 copies 2 mer at 73% conserved"

repeat\_region 12110..12177

repeat\_region /note="17 copies 4 mer atat 79% conserved"

repeat\_region 12790..12905

repeat\_region /note="L2 repeat: matches 2628. .2750 of consensus"

repeat\_region 13672..13923

repeat\_region /note="L2 repeat: matches 2462. .2725 of consensus"

repeat\_region 14774..15053

repeat\_region /note="AluSq repeat: matches 1. .289 of consensus"

repeat\_region 15966..16031

repeat\_region /note="6 copies 11 mer 77% conserved"

repeat\_region 17304..17359

repeat\_region /note="MIR repeat: matches 196. .251 of consensus"

repeat\_region 17554..17682

repeat\_region /note="MIR repeat: matches 108. .228 of consensus"

repeat\_region 18366..18701

repeat\_region /note="MER2 repeat: matches 1. .345 of consensus"

repeat\_region 18959..19248

repeat\_region /note="AluDo repeat: matches 1. .284 of consensus"

repeat\_region 20386..20592

repeat\_region /note="L2 repeat: matches 70. .288 of consensus"

repeat\_region 20609..20778

repeat\_region /note="LTR28 repeat: matches 848. .1013 of consensus"

repeat\_region 20747..21117

repeat\_region /note="LTR1 repeat: matches 295. .680 of consensus"

repeat\_region 21163..21226

repeat\_region /note="LTR28 repeat: matches 297. .361 of consensus"

repeat\_region 21187..21359

repeat\_region /note="LTR1 repeat: matches 16. .188 of consensus"

repeat\_region 22191..22355

repeat\_region /note="FRAM repeat: matches 4. .163 of consensus"

repeat\_region 22844..22991

repeat\_region /note="MIR repeat: matches 48. .202 of consensus"

repeat\_region 23155..23190

repeat\_region /note="18 copies 2 mer ga 86% conserved"

repeat\_region 24390..24440

repeat\_region /note="L2 repeat: matches 2657. .2708 of consensus"

repeat\_region 25082..25109

repeat\_region /note="7 copies 4 mer aac 100% conserved"

repeat\_region 25438..25744

repeat\_region /note="L2 repeat: matches 2427. .2744 of consensus"

repeat\_region 25751..25917

repeat\_region /note="MIR repeat: matches 95. .262 of consensus"

repeat\_region 25964..26307

repeat\_region /note="L2 repeat: matches 1986. .2366 of consensus"

repeat\_region 26526..26673

repeat\_region /note="FLAM C repeat: matches 1. .142 of consensus"

repeat\_region 26682..26737

repeat\_region /note="14 copies 4 mer agga 76% conserved"

repeat\_region 28208..28394

repeat\_region /note="MIR repeat: matches 12. .196 of consensus"

repeat\_region 29052..29250

repeat\_region /note="MIR repeat: matches 48. .251 of consensus"

repeat\_region 30082..30170

repeat\_region /note="MIR repeat: matches 50. .145 of consensus"

repeat\_region 30197..31090

/note="VER11D repeat: matches 1. .897 of consensus"  
31921. .32087  
/note="FRAM repeat: matches -2. .151 of consensus"  
32545. .32678  
/note="L2 repeat: matches 2578. .2708 of consensus"  
32860. .32895  
/note="9 copies 4 mer tttta 83% conserved"  
32897. .33193  
/note="AluX repeat: matches 1. .296 of consensus"  
35445. .35478  
/note="MIR repeat: matches 84. .118 of consensus"  
37442. .37598  
/note="VER5B repeat: matches 1. .173 of consensus"  
38659. .38860  
/note="MIR repeat: matches 15. .236 of consensus"  
40552. .40911  
/note="THE1B repeat: matches 1. .364 of consensus"  
40912. .42540  
/note="THE1B-INTERNAL repeat: matches 1. .1580 of  
consensus"  
42541. .42899  
/note="THE1B repeat: matches 1. .364 of consensus"  
42900. .43231  
/note="MT1A1 repeat: matches 4. .349 of consensus"  
43407. .43851  
/note="MT1C repeat: matches 9. .466 of consensus"  
44641. .45020  
/note="MSTA repeat: matches 1. .382 of consensus"  
45689. .45988  
/note="AluS repeat: matches 1. .302 of consensus"  
46220. .46351  
/note="Tigger3b repeat: matches 567. .698 of consensus"  
46358. .46609  
/note="AluJb repeat: matches 40. .290 of consensus"  
46623. .46759  
/note="AluJo/FRAM repeat: matches 170. .301 of consensus"  
46759. .47229  
/note="Tigger3b repeat: matches 2. .540 of consensus"  
47230. .47442  
/note="MER44A repeat: matches 85. .327 of consensus"  
47715. .48005  
/note="LIP1A10 repeat: matches 5864. .6153 of consensus"  
48577. .48973  
/note="MT1B repeat: matches 1. .390 of consensus"  
50633. .50925  
/note="AluY repeat: matches 3. .295 of consensus"  
53828. .55056  
/note="LIMC1 repeat: matches 5023. .6332 of consensus"  
55092. .56081  
/note="LIM4 repeat: matches 3778. .4798 of consensus"  
56074. .56651  
/note="LIMCA repeat: matches 579. .1150 of consensus"  
56652. .56952  
/note="AluSx repeat: matches 1. .301 of consensus"  
56953. .57277  
/note="LIMCA repeat: matches 253. .579 of consensus"  
57280. .57562  
/note="L2 repeat: matches 1590. .1886 of consensus"  
57699. .57766  
/note="L2 repeat: matches 2639. .2710 of consensus"  
59295. .59599  
/note="AluSx repeat: matches 1. .307 of consensus"  
59749. .59811

Alignment Scores:  
Pred. No.: 5 6e-63 Length: 129010  
Score: 67.00 Matches: 67  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.41% Indels: 0  
DB: 9 Gaps: 0

US-09-744-313a-1 (1-465) X AL589666 (1-129010)

Qy 56 IlesGluGluGlyIleValValMetGluAspAspSerProValGluAlaValSerThrPro 75  
Db 78112 ATTGAAGAAGGATTGTTGTAATGGAAGATGATCTCCAGTGAGGCTGTGAGCACCT 78053  
Qy 76 AsnThrProArgAsnLeuAlaLaTPlYsIleSerIleProTyrValAspPhePheGlu 95  
Db 78052 AATATCCCCGAACCTTGCTGATGGAATAATGATTCATATCCATATGATGATTTTGTAG 77993  
Qy 96 AspProSerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAspVal 115  
Db 77992 GATCCCTCTCTGAAAGGAAGGAGAAAAAGAAAGAAATTCCTGTGTTTGTATGTATGTT 77933  
Qy 116 GluArgAsnAspArgAla 122  
Db 77932 GAAAGAATGATAGAAGCA 77912  
RESULT 15  
AC116713/c  
LOCUS AC116713 278898 bp DNA linear HTG 09-JUN-2003  
DEFINITION Mus musculus clone RP23-118M23, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 7  
ordered pieces.  
ACCESSION AC116713  
VERSION AC116713.4 GI:131544102  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 278898)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP23-118M23  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 278898)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,J., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lechoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 278898)  
REFERENCE  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,

O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ranasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thoman, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Willson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (09-JUN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 9, 2003 this sequence version replaced gi:29171427.  
 All repeats were identified using RepeatMasker:  
 Smit, A.P.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center Project name: L22409  
 Center clone name: 118\_M\_23

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 97013: contig of 97013 bp in length  
 \* 97014 97113: gap of 100 bp  
 \* 97114 144931: contig of 47818 bp in length  
 \* 144932 145031: gap of 100 bp  
 \* 145032 171184: contig of 26153 bp in length  
 \* 171185 171284: gap of 100 bp  
 \* 171285 221380: contig of 50096 bp in length  
 \* 221381 221480: gap of 100 bp  
 \* 221481 224451: contig of 2971 bp in length  
 \* 224452 224551: gap of 100 bp  
 \* 224552 249724: contig of 25173 bp in length  
 \* 249725 249825: gap of 100 bp  
 \* 249825 278898: contig of 29074 bp in length.

#### FEATURES

Location/Qualifiers  
 1..278898  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-118M23"  
 /clone\_lib="RPCI-23 Female Mouse BAC"  
 BASE COUNT 84330 a 58884 c 57710 g 76354 t 1620 others

#### ORIGIN

Alignment Scores:  
 Pred. No.: 1..12e-62 Length: 278898  
 Score: 67.00 Matches: 67  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 14.41% Indels: 0  
 DB: 2 Gaps: 0

US-09-744-313A-1 (1-465) X AC116713 (1-278898)

QY 56 TleGluGluGlylleValValMetGluAspSerProValGluAlaValSerThrPro 75  
 Db 238593 ATTGAAGAGGGATTGTGTAATGAGGATGACTCTCCAGTAGAAGCTGTGAGCACACCT 238524  
 QY 76 AanthrProArgAsnLeuAlaAlaTtpylsileProTyrValAspPheGlu 95  
 Db 238523 AACACTCTCGAAACCTTGCTGCTCGGAAATAGCATTCGTATGATGAGACTTTTGTGAA 238464

QY 96 AspProSerSerGluArgLysGluLysGluArgLysGluArgLysGluArgLysGluArgLys 115  
 Db 238463 CATCCCTCCCTCTGAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 238404  
 QY 116 GluArgAsnAspArgArgAla 122  
 Db 238403 GAAAGAAACGACAGACAGAGCA 238383

Search completed: January 31, 2004, 08:16:12  
 Job time : 5396 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:44:36 ; Search time 85 Seconds

(without alignments)  
2414.625 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: MYLLSLCLFRNTQKRGESF.....ELFPELNKYQKEVTSWTSMW 465

Scoring table: BLOSUM62

	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09744313/runat 30012004 113536 4209/app query.fasta\_1.647  
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HRAPISE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313.@CIGN 1 1 56 @runat 30012004 113536 4209 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2386	97.9	1716	4	US-09-620-312D-922
2	2057.5	84.4	1551	4	US-09-620-312D-923
3	136.5	5.6	1497	4	US-09-620-312D-1021
4	128	5.3	1974	1	US-08-625-322-1
5	116	4.8	3747	2	US-09-080-897-1
6	116	4.8	3747	3	US-09-323-735-1
7	113.5	4.7	5822	3	US-08-899-595-4
8	113.5	4.7	5822	3	US-08-899-595-5
9	110	4.5	580073	4	US-08-545-528D-1
10	109.5	4.5	580073	4	US-08-545-528D-1
11	108.5	4.5	9409	4	US-08-961-527-161
12	108	4.4	3621	4	US-09-635-872A-21

13	108	4.4	3621	4	US-09-636-077A-21
14	108	4.4	3783	4	US-09-635-872A-20
15	108	4.4	3783	4	US-09-636-077A-20
16	105	4.3	4065	4	US-09-016-434-1105
17	105	4.3	4739	3	US-08-585-871-1
18	103	4.2	1353	1	US-08-625-322-3
19	103	4.2	5597	4	US-09-635-872A-4
20	103	4.2	5597	4	US-09-636-077A-4
21	103	4.2	1664976	4	US-08-916-421B-1
22	102.5	4.2	3095	6	5231168-1
23	102.5	4.2	4378	2	US-09-080-897-3
24	102.5	4.2	4378	3	US-09-323-735-3
25	102.5	4.2	4399	3	US-08-899-595-2
26	101.5	4.2	1330	2	US-09-036-582-33
27	101.5	4.2	8503	4	US-09-620-312D-130
28	101	4.1	5893	1	US-08-592-126-54
29	101	4.1	5893	2	US-08-687-080-44
30	101	4.1	5893	4	US-09-168-595-54
31	100.5	4.1	6519	1	US-08-588-985-1
32	100.5	4.1	6519	1	US-08-971-988-1
33	100	4.1	15223	3	US-08-962-690-12
34	100	4.1	15223	2	US-08-932-403A-1
35	100	4.1	15223	3	US-08-720-132-1
36	99.5	4.1	4100	4	US-09-620-312D-81
37	99	4.1	977	3	US-09-168-595-54
38	99	4.1	2027	4	US-09-484-970B-98
39	98.5	4.0	424	4	US-09-397-787-133
40	98.5	4.0	3309	4	US-09-510-543-20
41	98.5	4.0	4982	3	US-08-699-103B-1
42	98.5	4.0	4982	4	US-09-229-059-1
43	98.5	4.0	4982	4	US-09-628-133-1
44	98.5	4.0	5878	4	US-09-510-543-18
45	98.5	4.0	1664976	4	US-08-916-421B-1

#### ALIGNMENTS

#### RESULT 1

US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenchua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Reivan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: DT\_FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA



ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (71) .. (1543)  
 US-09-620-312D-922

Alignment Scores:  
 Pred. No.: 5,99e-281 Length: 1716  
 Score: 2386.00 Matches: 456  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.91% Indels: 0  
 DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
 Db 173 TTTCCGACACACACAGAAAGGGAGATCATTTGGATCAGCAGAAATAGGTAGCAAAATT 232  
 QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
 Db 233 AAAGGAGTATTCAAAGTACCAATGGAGGAGCTATGTTGCCTAAATTATGTTGAGCT 292  
 QY 50 GluGlyGluAspAspPheIleGluGlyIleValValMetGluAspAspSerProVal 69  
 Db 293 GAAGGTGAGATGATTTTATTGAAGAAGGTATTTGTTGAATGGAAGATGATCTCCAGTG 352  
 QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTTPLyIleSerIlePro 89  
 Db 353 GAGGCTGGACACACCTTAATCTCCCGAAACCTTGTGTCATGGAAATTTAGCATTTCCA 412  
 QY 90 TyrValAspPheGluAspProSerGluArgLysGluLysGluLysGluArgIlePro 109  
 Db 413 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAGGAGGAGGAAAGAAAGAAATTCCT 472  
 QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
 Db 473 GTGTTTGTATGTATGTTGAAGAATATGATAGAGCAGATTTGGACACGAGCCTGAACAT 532  
 QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
 Db 533 TGCTCTGTCTATAGAGATATCTTGAATTTATGATCTTGAATCAAAACTAACAGAAATT 592  
 QY 150 HisGlyValPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
 Db 593 CATGGTGATTTCTGATGCCAGCTTCTCTTAAGAGATCATTTGGCCCAAAATTTAT 652  
 QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
 Db 653 GAATTCCTTAAAGTCAAAAGGAGGAGAGATTTCCAGAAATATCTACAGAACTTCTGAGCAT 712  
 QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
 Db 713 CCAGAACTAGTAATAGTCAACTCTGGCAGACTTTCTTTCCTCCCTTAATGGTGGGAAACA 772  
 QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229  
 Db 773 CAATTTCTTGATAGATATCATCAGATGTAATCTTGGGAAATTTATAATCTGTTCT 832  
 QY 230 GlyLysLeuMetLysGlyLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
 Db 833 GGAATACTAATGAAGAGAAAGGTGAGCATTTTGGAACTTTTATCATGAATTTTCAATAT 892  
 QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
 Db 893 TCTTGTGAGTCTCAAGGCTTAACCAAGTAGACCACTGACCACTTCTCAGCCCTACT 952  
 QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
 Db 953 TCAGAAACACCAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 1012  
 QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309

Db 1013 AATACAGAGAGAAACAAATATTTATGAGGTGATGACTGTAGAGGAGTC 1072  
 QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
 Db 1073 TATGATTACCTCATGTATGTAGGACGGTAGTTTCCAGGTTCTGACTGGCTTCATCAT 1132  
 QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
 Db 1133 CTCCTTAATGGGAACTCGAATCTCTTTAAACACCCCTGGAAATGTATATCTGATTACTAT 1192  
 QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
 Db 1193 CTTCACTGTAAACTAGACAGCTATTTCCAGGAGCACCGTTTGGTCTCACTATAACACTT 1252  
 QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
 Db 1253 CTCAGAGATGCTATATTTCTGTGAAAAACATGAACTCGCTCTCTCCAAAGATAAGCAAAA 1312  
 QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuValLysCys 409  
 Db 1313 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTCATTCAGATCTGTTAGTCAAGTGT 1372  
 QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
 Db 1373 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCA 1432  
 QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
 Db 1433 GTACTCAACAAGCAGCTGACTTATGTTTATTTGGCACTGTGTATACAGGAACTGTTTCCA 1492  
 QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
 Db 1493 GAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 1540

# RESULT 2

US-09-620-312D-923  
 ; Sequence 923, Application US/09620312D  
 ; Patent No. 659662  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Asundi, Vinod  
 ; APPLICANT: Zhang, Jie  
 ; APPLICANT: Ren, Feiyan  
 ; APPLICANT: Chen, Rui-hong  
 ; APPLICANT: Zhao, Qing A.  
 ; APPLICANT: Wehrman, Tom  
 ; APPLICANT: Xue, Aidong J.  
 ; APPLICANT: Yang, Yonghong  
 ; APPLICANT: Wang, Jian-Rui  
 ; APPLICANT: Zhou, Ping  
 ; APPLICANT: Ma, Yunding  
 ; APPLICANT: Wang, Dunrui  
 ; APPLICANT: Wang, Zhiwei  
 ; APPLICANT: John Tillinghaast  
 ; APPLICANT: Drmanac, Radoje T.  
 ; TITLE OF INVENTION: No. 659662el Nucleic Acids and  
 ; TITLE OF INVENTION: Polypeptides  
 ; FILE REFERENCE: 784CIP2B  
 ; CURRENT APPLICATION NUMBER: US/09/620,312D  
 ; CURRENT FILING DATE: 2000-07-19  
 ; PRIOR APPLICATION NUMBER: 09/552,317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: 09/488,725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; NUMBER OF SEQ ID NOS: 1105  
 ; SOFTWARE: pt FL\_genes Version 1.0  
 ; SEQ ID NO 923  
 ; LENGTH: 1551  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:

NAME/KEY: CDS  
LOCATION: (71)..(1378)  
US-09-620-312D-923

Alignment Scores:  
Pred. No.: 6,28e-241 Length: 1551  
Score: 2057.50 Matches: 401  
Percent Similarity: 87.94% Conservative: 0  
Best Local Similarity: 87.94% Mismatches: 0  
Query Match: 84.43% Indels: 55  
DB: 4 Gaps: 1

US-09-744-313a-1 (1-465) x US-09-620-312D-923 (1-1551)

QY 10 PheArgAsnThrGlnLysArgGlyGlySerPheGlyLeuSerArgGlySerLysLeu 29  
DB 173 TTTCCGACACACAGAAAGGGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATT 232  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyValAlaMetLeuProAsnTyrGlyValAla 49  
DB 233 AAAGGATATTCAAAGATACCAATGAGGGAGCTATGTTGCTAATTATGTTAGCT 292  
QY 50 GluGlyGluAspPheIleGluGlyValValMetGluAspSerProVal 69  
DB 293 GAAGGTGAAGATGATTTTATTAAGAAGGATTTGTTGTAATGGAAGATGATTTCCAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaIleLysIleSerIlePro 89  
DB 353 GAGGCTGTGACACACACCTTAATACCTCCCGAAACCTTCTGCATGGAAATTAGCATTTCCA 412  
QY 90 TyrValAspPheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro 109  
DB 413 TATGTAGACTTTTGTGAGATCCCTCTCTCTGAAGAGAGAGAGAGAGAGAGAGATTCCT 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
DB 473 GTGTTTCTGATGTTGTTGAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGATTTAT 532  
QY 130 TrpSerValTyrArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 533 TGGTCTGTCTAGAGATATCTTGAATCTTATGATCTTGAATCAAACTTAACAGAGATTT 592  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
DB 593 CATGTTGATTTCTCTGATGCTCCCTCTTCTAAGAGAGATCATTTGCCCCCAAAATTTAT 652  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
DB 653 GAATTTCTTAAGTCAAGAGAGAGAGAGTTCAGAGATATCTACAGAACTTCTGCAGCAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
DB 713 CCAGAACTGAGTAAATAGTCAACTTCTGGCAGACTTTCTTTCCCTAATGTTGGGAAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
DB 773 CAATTTCTTTGATAGATACTACAGATGTAATCTTGGGAAATTAATAAATCTGTCTCT 832  
QY 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
DB 833 GGAAACTTAATGAAGAGAAAGGTGAGCATTTTGAACCTTTATCATGAATTTTCATTAAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
DB 893 TCTTGTGATGTTCCAAAGCTTAACCAAGTAGAGACAGACTGACCATCTTCAGCCCTACT 952  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
DB 953 TCAGAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
DB 1013 AATACAGAGAGAGAGAGAGAGAGATTTATTTTATGGAGGTGATGACTGTAGAGGATC 1072

## RESULT 3

US-09-620-312D-1021  
Sequence 1021, Application US/09620312D  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Wehrman, Tom  
APPLICANT: Zhao, Qing A.  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yungqing  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP28  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt\_fl\_genes Version 1.0  
SEQ ID NO 1021  
LENGTH: 1497  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (79)..(762)

QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
DB 1073 TATGATTACCTGATGATGTA----- 1093  
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1093 ----- 1093  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
DB 1093 ----- 1093  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 1094 -----GATGTTATATCTCTGAAACCACTGAACTCGCTCTCTCCAGATAAGCAAAA 1147  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
DB 1148 GGAGCAAAACAGACTTTTGAAGAAATGATGAATACATTCAGATCTCTTGTAGTCAAGTGT 1207  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
DB 1208 ATTTGTTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGTATGGCTTACAGCAACA 1267  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
DB 1268 GTACTCAACAGACACTGACTTATGTTTATTTGACATTTGATGATACAGAACTGTTTCCA 1327  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
DB 1328 GAGCTCAATAAGGTACAAAGAGAGTACCTCTGTGACATCTTGGATG 1375

## US-09-620-312D-1021

Alignment Scores:  
Pred. No.: 1.71e-06 Length: 1497  
Score: 136.50 Matches: 63  
Percent Similarity: 39.63% Conservative: 44  
Best Local Similarity: 23.33% Mismatches: 108  
Query Match: 5.60% Indels: 55  
DB: 10 Gaps: 10

## US-09-744-313A-1 (1-465) x US-09-620-312D-1021 (1-1497)

QY 68 ProValGluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSer 87  
DB 25 CCGGTGAATGGGAAGATAGACCACTACCTACTACTACTG----- 66  
QY 88 IleProTyrValAspPheGluAspProSerSerGluArgLysGluLysLysGluArg 107  
DB 67 -----GGTATGAAGTATGATGAAGAAAGAGCTAAA 96  
QY 108 IleProValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluPro 127  
DB 97 TTTACTGTATATAAATACTAGTAAGAAACCCAGAA----- 135  
QY 128 GluHisTrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThr 147  
DB 136 GAAAGCTGGGTATTTTCAGAGATACACTGACTTCTCTAGCTTAATGACAAATTAATA 195  
QY 148 GluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleGlyProLys 167  
DB 196 GAGATGTTCCAGTTTCGACTACACTTCTCCAAACCGCTGTTTAAAGTAATATAC 255  
QY 168 AsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGluLysLeuLeu 187  
DB 256 AATGCTGACTTTTGAAGACAGACAAATTAGGATTACAAGCGTTCTTCAAAATTTAGTA 315  
QY 188 GlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeu-----SerPro 204  
DB 316 GCTCACAAGGACATTGCTAACTGCTTGCAGTGCAGAGAAATTTCTTTTGGATGATCCA 375  
QY 205 AsnGly-----GlyGluThrGlnPheLeuAspLysIleLeuProAspVal 219  
DB 376 CCGGTGCCATTTGATAGCTAGAGAAAGACGGGCAATCTGTGAACCTTTAGAGAGACA 435  
QY 220 AsnLeu-----GlyLysIleIleLysSerValProGly 230  
DB 435 AACTACCGCTTACAGAAAGAACTACTTGAAACAAAGAGAGATGGAATCACTAAAG-- 492  
QY 231 LysLeuMetLysGluLysGlyGlnHisLeuGluPheIleMetAsnPheIleAsnSer 250  
DB 493 AACTGCTCAGTGAGAAGCACTTCATATAGACACT--TTAGAGACAGAAATCAGAAC 549  
QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
DB 550 TTG---TCTTTAGAACCTGAAGATCA-----CTGGATGTCTCAGAACAGAGGT 597  
QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
DB 598 GAACAGATCTTAAGGTGGAGCTCTCTGCACCTTGAGTTGATCAAGATGCTCTGATGAA 657  
QY 291 ThrGluArgLysGlnAsn-----GlnAsnTyrPheMet 301  
DB 658 GAATCTAGAGCTGATATAAACCATGCTTAAGTTTGTAGTGAACCTGAAATGCTGTATCA 717  
QY 302 GluValMetThrValGluGlyValTyrAsp 311  
DB 718 GAGATAGAAGTAGCAGAGAGTGGCATATGAT 747

## RESULT 4

US-08-625-322-1  
; Sequence 1, Application US/08625322  
; Patent No. 5804412  
; GENERAL INFORMATION:

APPLICANT: Gill, Gordon N.  
APPLICANT: Kurten, Richard C.  
APPLICANT: Cadena, Deborah L.  
TITLE OF INVENTION: Sorting Nexins and Methods of Using Same  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/625,322  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UD 1955  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1974 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45..1612  
US-08-625-322-1

## Alignment Scores:

Pred. No.: 2.96e-05 Length: 1974  
Score: 128.00 Matches: 98  
Percent Similarity: 34.79% Conservative: 69  
Best Local Similarity: 20.42% Mismatches: 187  
Query Match: 5.25% Indels: 126  
DB: 19 Gaps: 19

## US-09-744-313A-1 (1-465) x US-08-625-322-1 (1-1974)

QY 26 GlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuPro--- 44  
DB 190 GGGCGCGGTGGTCAAGTAAACATCATGCTCCAAAGATAACTACATCCCTTCTTCCATC 249  
QY 45 -----AsnTyrGlyValAlaGluGlyGluAsp-----AspPhe 55  
DB 250 AACAAATGGTCCAAAGAAATGGATCCATGAGAACCAAGACCAAGAGCCAGAGATCTC 309  
QY 56 IleGluGluGlyIleValValMetGluAspSerProValGluAlaValSerThrPro 75  
DB 310 TTTGCAGATCCACAGTGGAGCTATCTCTGGACAGC-----ACACAAAT 354  
QY 76 AsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPheGlu 95  
DB 355 AATCAGAAGAAGGTGCTAGCCAAACACACTATTTCCTCTCCTCAGGAAGCCACAAAT 414  
QY 96 -----AspProSerSerGluArgLysGluLysGluArgIleProValPhe 111  
DB 415 TCTTCGAAGCCCGAGCCAACTATGAGGAGCTAGAGGAGAGAAAGAACAGAGGATCAATTT 474  
QY 112 CysIleAspValGluArgAsnAspArgAlaValGly----- 124  
DB 475 GATTTGACATCGGTATACTGATGATGAGAGATAGGAGGATGATGATGATGATGATGTA 534

125 -----HisGluProGluHisTrpSerVal 132  
535 GCTTACAAAGTTTACACACAGACAGCTTACCATTTTTCAGAAAGCAACAGCTTTCAGTA 594  
133 TyrArgGlyTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis----- 150  
595 AAAAGAGAGATTAGTGTCTTCTGGTCTTTATGAGAGCTTTCCGAGAGACACTTCTCAG 654  
151 GlyAlaPheProAspAlaGlnLeuProSerLysArgGlyLeuGly----- 165  
655 AATGGCTTATGTCCTCCGCGCCCGGAGAGAGCCTCATAGGATGCAAAAGTGAA 714  
166 -----ProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGln 180  
715 GTTGGAGAGAGATTCTTCTCTGCAAGATTCTTGAAGAGCGGCGCTTTAGAA 774  
181 GluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
775 AGTACCTTCAGAGGATTGTAATCATCTTACATCTTACAGGACCTGACGCTCAGAG 834  
201 PheLeuSer-----ProAsnGlyGlyGluThrGlnPheLeu----- 212  
835 TTTTGGAAAGAGAGAGTGCACGTCGCGTGGGTACCCACACATTGAGTGTCTGCT 894  
213 -----AspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSer 227  
895 CTCTCAGATGTTCAACAGCCACAGATGCGTCCAGCAAAATGACCATCAAGATGAAT 954  
228 ValProGlyLysLeuMetLysGlnHisLeuGluProPheIleMetAsnPhe 247  
955 GAATCAGACATTGTTTGGAGAGAGCTCCAGGAGTAGAG----- 996  
248 IleAsnSerCysGluSerProLysProLysPro----- 258  
997 -----TGTAGAGAGAGCGCTTACGGAACCTGCTGTGTGTAGAAACTTCTAGTC 1047  
259 ---SerArgProGluLeuThrIle-----LeuSer 267  
1048 AACCATAGGAAGAGTGTGCTGACACAGCCAGCTTTCAGAGAGTACCATGCTT 1107  
268 ProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu-----PheLysAsnAsnAla 285  
1108 GCGAGCTCTGAGGACACAGCGCATTTGTACCGGGCACTCTCCAGCTGCTGAGTGGAA 1167  
286 AsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThr 305  
1168 GAAAAAATTGAGCAGCTCCACAGGAGCCCAACAAATGACTTCTCTCTCTCTGCTGAG 1227  
306 ValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPhe-----GlnVal 323  
1228 CTCTGAGTACTACATTGCTCTCTGCGCATAGTCCGCTGCTGCTGCTGCTGCTGCTGCT 1287  
324 ProAspTrpLeuHis-----HisLeuLeuMetGly 333  
1288 AAGCATGGCAGCGCTGCGAGGATGCCAGCCACACTGCAGAGAGGAGGCGGAG 1347  
334 ThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLys 353  
1348 GCTCGGCTGTGTGGGCAACAG-----CCTGATAAGCTGCGAGCGGCAAG 1395  
354 LeuGluGlnLeuPheGlnGluHisArgLeuValSer-----Leu 366  
1396 GAGCAGATCTCGAGTGGGAGTCTCGGTGCTGCTCATATGAAGAGGACTTCGAGAGGATT 1455  
367 IleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAsp 386  
1456 TCACAGTGTGCGAAAGAGAGTATA-----CGGTTGGAAGAGAG 1497  
387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
1498 AATCCAGAGACTTCAAG-----AACCAAGTGTACAGTACCTTGTGAGACACTCTT 1548

RESULT 5

US-09-080-897-1  
Sequence 1, Application US/09080897  
Patent No. 5985574  
GENERAL INFORMATION:  
APPLICANT: King, Mary-Claire  
APPLICANT: Lynch, Eric D.  
APPLICANT: Lee, Ming  
APPLICANT: Morrow, Jan E.  
APPLICANT: Welch, Piri L.  
APPLICANT: Leon, Pedro E.  
TITLE OF INVENTION: Modulators of Actin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 75 DENISE DRIVE  
CITY: HILLSBOROUGH  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/080,897  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UW97-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 343-4341  
TELEFAX: (650) 343-4342  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3747 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1...3744  
US-09-080-897-1  
Alignment Scores:  
Pred. No.: 0.00252 Length: 3747  
Score: 116.00 Matches: 79  
Percent Similarity: 40.10% Conservative: 75  
Best Local Similarity: 20.57% Mismatches: 134  
Query Match: 4.76% Indels: 96  
DB: 2 Gaps: 20  
US-09-744-313A-1 (1-465) x US-09-080-897-1 (1-3747)  
QY 98 SerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArg 117  
DB 2395 TCTCCAGACCAAGACCAAGAGGATCAA----- 2424  
QY 118 AsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArgGlyTyrLeu 137  
DB 2425 -----GAAGGTGGAGAAAGAAATCTCTGCAAAAGAAAAGTAAAA 2469  
QY 138 GluPheTyrValLeuGluSerLys-----LeuThrGluPheHisGlyAlaPhe 153  
DB 2470 GAGTAAAGGTGTGGATTCAAGACAGCCAGCAATCTCTCAATCTTTTGGGTCTCTTC 2529  
QY 154 ProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyrGluPheLeuLys 173  
DB 2530 -----CGCATGCGCTATCAAGAGATT-----AAGATGTCTATCCTCGAGGTG 2571



2632 GAGCAGTAAAGATCTTCTGAAGTGAAGAGAA-----TAT 2670  
213 AspLysLeuProAspValAsnLeuGlyLysLeuValProGly----- 230  
2671 GATGACCTGGTGAAGTCAAGAGATTTGGCGTGGTGAAGGACATGTCGCCGACTGGG 2730  
231 -----LysLeu---MetLysGluLysGlyGlnHisLeuGlu 241  
2731 CTTGGCTCAATGCCATTCTCTCAAGCTACAATTCAAGGAGCAAGTGAAGATATCAAG 2790  
242 ProPheLeuMetAsnPheLeuLeuSerCysGluSerProLysProLysProSerArgPro 261  
2791 CAGAGATTTGTTCTCTACTGCTCATGCTGAGGAGTTTCAAGAGTGAAGGAGTTCCTTCC 2850  
262 -----GluLeuThrLeu-----SerProThrSerGluAsn 273  
2851 ATCTCTAGAGATTAACCTTCTGTTGGAAATTCATGATGCTGGCTCCAGAAATGCT 2910  
274 LysLeuLeu---PheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGlu 292  
2911 GGTGCTTTGGCTTCAATATCAGCTTCTCTGTAAGCTTCGAGACACCAAGTCCACAGAT 2970  
293 ArgLysGluAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyr 312  
2971 CAGAGATGAGCTGTTTACACTTCTGCTGAGTGTGAGAGTGAAGTATCCCATGTC 3030  
313 LeuMetTyrValGlyArgValPheGlnValPheGlnValProAspTyrPheHisLeuLeuMet 332  
3031 CTAAGTTT-----CCAGACGAGCTTGGCCATGTGGAGAA 3066  
333 GlyThrArgLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCys 352  
3067 GCAGCGGAGTTCTCTGAAAC-----TTGCAAAAG 3099  
353 LysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArgAsp 372  
3100 AACCTAGATCAGATGAAGAACAA-----ATTCTGATGTGAAGCTGAT 3144  
373 AlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLys 392  
3145 GTT-----CAGAAATTTCCAGCTGCCACAGATGAAGAAAGACAAAG----- 3183  
393 GlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGlu 412  
3184 ---TTGTGTAAGAAATGACCACTTGTGAAGAT-----GCA 3219  
413 GluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsn 432  
3220 CAGAACAGTATTAACAGCTGGGATGATGATCTTAACATGAG-----ACCTCTAT 3273  
433 LysGlnLeu---ThrTyrValLeuLeuAsp-----IleValIleGlnLeuPhe 448  
3274 AAGGAGCTGGCGAGTACTTCTCTTGTACCCCAAGAGTGTCTGTGAAGAAATTTTC 3333  
449 ProGluLeuAsn 452  
3334 ATGGATCTTCAC 3345

STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/899,595  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-242701  
FILING DATE: 26-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 9-90170  
FILING DATE: 25-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stephen A. Bent  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 049441/0112  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5822 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 28..3972  
US-08-899-595-4

Alignment Scores:  
Pred. No.: 0.0106 Length: 5822  
Score: 113.50 Matches: 73  
Percent Similarity: 41.09% Conservative: 70  
Best Local Similarity: 20.98% Mismatches: 124  
Query Match: 4.66% Indels: 81  
DB: 3 Gaps: 19

US-09-744-313A-1 (1-465) x US-08-899-595-4 (1-5822)

Qy 134 ArgArgTyrLeuGluPheTyrValLeuGluSerLys-----LeuThrGluPhe 149  
Db 2686 AAAAAAGTAAAGAGTTTAAAGGTTTGGATTCAAGACAGCCAGCAATCTCTCAATCTTT 2745  
Qy 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 2746 TTGGTTCTCTTC-----CCATGCCCTATCAAGATT-----AGATGTC 2787  
Qy 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeu---Gln 188  
Db 2788 ATCTCGAGGTGAATGAGGCTGTTCTGACTGACTATGATCCAGAACCTCATTAAGCAA 2847  
Qy 189 HisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGlu 208  
Db 2848 ATGCCAGAGCCAGAGCAGTATAAATGCTTCTCAACTGAAGATGAA----- 2895  
Qy 209 ThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerVal 228  
Db 2896 -----TATGATGACCTGGCTGAGTTCAGAGCAGTTTGGCTGGTGGTGGCAGCTGTG 2946  
Qy 229 ProGly-----LysLeu---MetLysGluLysGly 237  
Db 2947 CCCCAGCTGGGGCTCGCCTCAATGCCATTCTCTCAAGCTACATTCAGCGAGCAAGTG 3006

RESULT 7  
US-08-899-595-4  
Sequence #4, Application US/08899595  
Patent No. 6111072  
GENERAL INFORMATION:  
APPLICANT: Narumiya, Shuh  
APPLICANT: Takahashi, No. 6111072uaki  
TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington

QY 238 GlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLys 257  
Db 3007 GAGATATCAAGCCAGAGATGTGTCTGCTCAGCTGCATGTGAGGATTCACAGAGT 3066  
QY 258 ProSerArgPro-----GluLeuThrIleLeu-----SerProThr 269  
Db 3067 GAGAGCTTTTCCATCTCTAGAGATTACCTTGTCTGAAATACATGAATGCTGGC 3126  
QY 270 SerGluAsnAsnLysLysLeu---PheAsnAspLeuPheLysAsnAsnAlaAsnArgAla 288  
Db 3127 TCCAGAAATGCTGTGCTTTGGCTTCAATATCAGCTTCTCTGTAAGCTTCGAGACACC 3186  
QY 289 GluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGly 308  
Db 3187 AAGTCCACAGATCAGAGATCAGCTTGTACCTTCTGCTGAGTGTGTGAGATGAC 3246  
QY 309 ValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLysHis 328  
Db 3247 TATCCGATGCTCTCAAGTTT-----CCAGACGAGCTTGGC 3282  
QY 329 HisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyr 348  
Db 3283 CATGTGGAGAAAGCCAGCGAGTTTCTGCTGAAAC----- 3318  
QY 349 TyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThr 368  
Db 3319 ---TTGCAAAAGAACCTAGATCAGATGAAGAACAA-----ATTCTGAT 3360  
QY 369 LeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGln 388  
Db 3361 GTGGAGCTGATGTT-----CAGAAATTTCCAGCTGCCACAGATGAAGAAC 3408  
QY 389 LysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLys 408  
Db 3409 AAG-----TTGTGTAATAAATGACACAGCTTTGTGAGGAT----- 3444  
QY 409 CysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGln 428  
Db 3445 ---GCACAGGACAGATTAACAGCTGCGGATGATGCATTTCAATGAGGAG--- 3492  
QY 429 ProValLeuAsnLysGlnLeu---ThrTyrValLeuLeuAsp-----IleValIle 444  
Db 3493 ---ACCTCTATAGAGCTGGCGAGTACTTCTCTTTGACCCCAAGAGTGTCTGTT 3549  
QY 445 GlnGluLeuPheProGluLeuAsn 452  
Db 3550 GAAGAAATTTTCATGATCTTCAC 3573

## RESULT 8

US-08-899-595-5/c

Sequence 5, Application US/08899595

Patent No. 6111072

GENERAL INFORMATION:

APPLICANT: Narumiya, Shuh

APPLICANT: Takahashi, No. 6111072uaki

TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESS: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,595

FILING DATE: 24-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-242701  
FILING DATE: 26-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 9-90170  
FILING DATE: 25-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stephen A. Bent  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 049441/0112  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5922 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: Human  
US-08-899-595-5

## Alignment Scores:

Pred. No.: 0.0106 Length: 5922  
Score: 113.50 Matches: 73  
Percent Similarity: 41.09% Conservative: 70  
Best Local Similarity: 20.98% Mismatches: 124  
Query Match: 4.66% Indels: 81  
DB: 3 Gaps: 19

US-09-744-313a-1 (1-465) x US-08-899-595-5 (1-5922)

QY 134 ArgArgTyrLeuGluPheTyrValLeuGluSerLys-----LeuThrGluPhe 149  
Db 3137 AAAAAGTAAAGAGTTTAAAGTGTGTGATTCAAGACAGCCAGATCTCTCAATCTTT 3078  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyr 169  
Db 3077 TTGGGTCTCTTC-----CGCATGCCCTATCAGAGATT-----AAGATGTC 3036  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnTyrLeuGlnLysLeuLeu---Gln 188  
Db 3035 ATCTGGAGGTGATGAGCTGTTCTGACTGAGTCTATGATCCAGAACCTCATTAAAGCAA 2976  
QY 189 HisProGluLeuSerAsnSerGlnLeuLeuAlaPheLeuSerProAsnGlyGlyGlu 208  
Db 2975 ATGCCAGAGCCAGAGCAGTAAATAATGCTTTCTGAACCTGAAGGATGAA----- 2928  
QY 209 ThrGlnPheLeuAspLysIleLeuProAspValaLeuGlyLysIleIleLysSerVal 228  
Db 2927 -----TATGATGACCTGGCTGAGTCAGAGCAGTTTGGCGTGTGATGGGACTGTG 2877  
QY 229 ProGly-----LysLeu---MetLysGluLysGly 237  
Db 2876 CCCCAGCTCGGCCTCGCTCAATGCCATTCTCTCAAGCTACAATTCAAGCAGCAAGTG 2817  
QY 238 GlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLys 257  
Db 2816 GAGATATCAGCCAGAGATGTGTCTGCTCACTGTGTCATGTGAGAGTTACGTAAGAGT 2757  
QY 258 ProSerArgPro-----GluLeuThrIleLeu-----SerProThr 269  
Db 2756 GAGAGCTTTTCCAATCTCTAGAGATTACCTTGTGTTGTAATAATCATGAATCTCGC 2697  
QY 270 SerGluAsnAsnLysLysLeu---PheAsnAspLeuPheLysAsnAsnAlaAsnArgAla 288  
Db 2696 TCCAGAAATGCTGGTGTCTTTGGCTTCAATATCAGCTTCTCTGTGAAGCTTCGAGACACC 2637



QY 289 GluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGly 308  
 Db 2636 AAGTCCACAGATCAGAGATGAGTGTGTACACTTCTTGCTGAGTGTGTGAGATGAC 2577  
 QY 309 ValTyrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrpLeuHis 328  
 Db 2576 TATCCCGATGTCCTCAAGTTT-----CCAGACGAGCTTGCC 2541  
 QY 329 HisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyr 348  
 Db 2540 CATGTGAGAAACCCAGCGAGTTCTGCTGAAAC----- 2505  
 QY 349 TyrLeuGlnCysLysLeuGlnLeuPheGlnGlnHisArgLeuValSerLeuIleThr 368  
 Db 2504 ---TTGCAAAAGAACCTAGATCAGATGAGAAACAA-----ATTCTGAT 2463  
 QY 369 LeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGln 388  
 Db 2462 GTGGAACGTGATGTT-----CAGAATTTCCAGCTGCCACAGATGAAAGAGAC 2415  
 QY 389 LysGlyAlaLysGlnThrPheGlnGluMetMetAsnTyrIleProAspLeuValLys 408  
 Db 2414 AAG-----TTTGTGAAAAAATCACCAGCTTTGTGAAGGAT----- 2379  
 QY 409 CysIleGlyGluThrLysTyrGluSerIleArgLeuPheAspGlyLeuGln 428  
 Db 2378 -----GCACAGAACAGTATAACAGCTCGGATGATGATCTTAACATGGAG--- 2331  
 QY 429 ProValLeuAsnLysGlnLeu---ThyTyrValLeuLeuAsp-----IleValIle 444  
 Db 2330 ---ACCTCTATAAGAGCTGGCGAGTACTTCTTTGACCCCAAGAGTTGTCGT 2274  
 QY 445 GlnGluLeuPheProGluLeuAsn 452  
 Db 2273 GAAGAAATTTTCATGGATCTTCAC 2250  
 RESULT 9  
 US-08-545-528D-1  
 ; Sequence 1, Application US/08545528D  
 ; Patent No. 6537773  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fraser et al.  
 ; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment  
 ; Patent No. 6537773  
 ; FILE OF INVENTION: Thereof, and Uses Thereof  
 ; FILE REFERENCE: PB193P1  
 ; CURRENT APPLICATION NUMBER: US/08/545,528D  
 ; CURRENT FILING DATE: 1995-10-19  
 ; PRIOR APPLICATION NUMBER: US 08/488,018  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: US 08/473,545  
 ; PRIOR FILING DATE: 1995-06-07  
 ; NUMBER OF SEQ ID NOS: 1  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 580073  
 ; TYPE: DNA  
 ; ORGANISM: Mycoplasma genitalium  
 US-08-545-528D-1  
 Alignment Scores:  
 Pred. No.: 60.9 Length: 580073  
 Score: 110.00 Matches: 94  
 Percent Similarity: 35.66% Conservative: 80  
 Best Local Similarity: 19.26% Mismatches: 152  
 Query Match: 4.51% Indels: 162  
 DB: 4 Gaps: 21  
 US-09-744-313A-1 (1-465) x US-08-545-528D-1 (1-580073)  
 QY 53 AspAspPheIleGlnGluGlyIleValMetGluAspSerProValGluAlaVal 72  
 |||:|||||

Db 259014 GATGAATTTTATGAGAACTACATT----- 259037  
 QY 73 SerThrProAsnLeuAlaIleTrpLysIleSerIle----- 88  
 |||:|||||  
 Db 259038 ---ACCCCTCAACCCCAAGTTCAATACTACCTCAACCTATCCATACATAGACCTTA 259094  
 |||:|||||  
 QY 89 -----ProTyrVal----- 91  
 |||:|||||  
 Db 259095 CTATGATGAACCTATTATACGCTGGGTTTAGAAGAGGGGTTACCGTGATGACTTCTATGA 259154  
 |||:|||||  
 QY 92 -----AspPhePheGluAspProSerSerGluArgLys 102  
 |||:|||||  
 Db 259155 ATAAAAACCGTGTTTTAACTAATGAACCACTTTGATAAAAAACCTTCGCTGCAA----- 259208  
 |||:|||||  
 QY 103 GluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArgAla 122  
 |||:|||||  
 Db 259209 -----CCAAATTTAT-----GACATTTGGTTTGTATGAT----- 259235  
 |||:|||||  
 QY 123 ValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyrValLeu 142  
 |||:|||||  
 Db 259236 ---GGTTATCTCCAAAGTGAATGAAAAAATCGTTCTTAAACCGGATGTTGATAAGATC 259292  
 |||:|||||  
 QY 143 GluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArg 162  
 |||:|||||  
 Db 259293 GAAAAACCGCTTTTAAAGAGATTTAAAGCGCTGGAAGATGAACCT----- 259337  
 |||:|||||  
 QY 163 IleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluPheGlnGluTyr 182  
 |||:|||||  
 Db 259338 -----AAAAACCTTAAAGCGCTTGAAGATGATTAACCGGTTGATTAAGTAA 259382  
 |||:|||||  
 QY 183 LeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeu 202  
 |||:|||||  
 Db 259383 CTTGATAAAAAGATTACCACTTGAAGTTGATTAACCGGTTGTTGTTATGATATAA 259442  
 |||:|||||  
 QY 203 SerProAsnGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGly 222  
 |||:|||||  
 Db 259443 AACTTCCAGTTCCAAAGAACCAACATCGTTGATAAGGTTAGTGAACCTGATAACTTAACC 259502  
 |||:|||||  
 QY 223 LysIleIleLysSerValProGlyLysLeuMetLysGlu----- 235  
 |||:|||||  
 Db 259503 CGTTTTTATAGATGAATGAACCAACCGCTTACAAAGAAACCGTGATTTTCTCAACTCC 259562  
 |||:|||||  
 QY 236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer----- 250  
 |||:|||||  
 Db 259563 AAGTATGCTAATTTAGTAACTTCCAAAGTAACTACCACTAACTAAATGATTTTCAC 259622  
 |||:|||||  
 QY 251 -----CysGluSerProLysProLysProSerArg----- 260  
 |||:|||||  
 Db 259623 CGCTTAATAGAAATCAAAACCAACCAATTAACCGCTTAAACCAAGATTAATGGTAAC 259682  
 |||:|||||  
 QY 261 -----ProGluLeuThrIleLeuSer 267  
 |||:|||||  
 Db 259683 CAAAATCTGATTGATAATAACGTTGCTTTACTGCAAAACCCCAACACACACAGTT----- 259736  
 |||:|||||  
 QY 268 ProThrSerGluAsnAsnLysLysLeuPheAsn-----AspLeuPheLysAsnAsnAla 285  
 |||:|||||  
 Db 259737 -----GAAAAAAGAACTACTTACTAAATGTTATTGATCAACTTTTCAATGAGCTT 259787  
 |||:|||||  
 QY 286 AsnArgAlaGluAsnThrGluArgLys-----GlnAsnGlnAsnTyrPheMetGlu 302  
 |||:|||||  
 Db 259788 GATCAACTTGAGAAATCAAAAGATTAATTAAGTATTGATGATGAATACTATAGAGAG 259847  
 |||:|||||  
 QY 303 ValMetThr-----ValGluGlyValTyrAspTyrLeuMetTyrValGlyArg 318  
 |||:|||||  
 Db 259848 TTAGTTAGTCAGATTAATGAATGCAAAATGTTTATGAAAAACATCGATCAA----- 259898  
 |||:|||||  
 QY 319 ValValPheGlnValProAspTrpLeuHisIleLeuMetGlyThrArgIleLeuPhe 338  
 |||:|||||  
 Db 259899 -----AATCAGATCCAGTTT 259913  
 |||:|||||  
 QY 339 LysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPhe 358  
 |||:|||||  
 Db 259914 AAACCAACCACTTATAGAGATGAGTTAAGTCAACTTGAGCGCAAGATCCAGCTC 259973  
 |||:|||||



QY 434 GlnLeuThrTyrrValleuLeuAspIleValleuGlnLeuPheProGluLeuAsnLys 453  
Db 411445 CACTCG-----CAAGCAAAATTCGAGAGTTAGAGAA 411413  
QY 454 ValGlnLysGluValThrSerValThrSer 463  
Db 411412 TGGAAATGAAGAGAGAGTAATTAATACC 411383

RESULT 11  
US-08-961-527-161  
Sequence 161, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 161:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9409 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-161

Alignment Scores:  
Pred. No.: 0.0961 Length: 9409  
Score: 108.50 Matches: 112  
Percent Similarity: 34.13% Conservative: 73  
Best Local Similarity: 20.66% Mismatches: 140  
Query Match: 4.45% Indels: 217  
DB: 4 Gaps: 31

US-09-744-313A-1 (1-465) x US-08-961-527-161 (1-9409)  
QY 25 IleGlySerLysIleLysGlyValPheLysSerThrMetGluGluValMetLeuPro 44  
Db 2724 TTGGATGCTCATATCAAGAACTGACAGGATTCAGACCAACGCTCTGGCGAAGCACCT 2783  
QY 45 AsnTyrrGly---ValalaGluGlyGluAspPheIleGluGluGlyValVal--- 62  
Db 2784 GATTTTCCAGATTGCCAGAAAATATTGACTTGGTGGAGATGGGATTTTGTAGCC 2843  
QY 63 -----MetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80  
Db 2844 CATAATGTTTCAGTTTGTATCTCTTGGCG-----GAAAAT 2882

QY 81 Leu-----AlaAlaTrpLysIleSerIleProTyrrValasp----- 92  
Db 2883 TTATTTTGAAGGCTATAGCTAAAGAACCTCTGTTGTATACGTCGTAATGGCCAG 2942  
QY 93 ---PhePheGluAspProSerSerGluArgLysGluLysGluArgIleProValPhe 111  
Db 2943 GTCTTTTC-----CCTGAACGGAA-----AAATATAGCTTGGCAATTTG 2984  
QY 112 CysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSer 131  
Db 2985 TGT-----CGAGAATTAGGAATTCCTCTTAACAC----- 3014  
QY 132 ValTyrrArgTyrrLeuGluPheTyrrValLeuGluSerLysLeuThrGluPheHisGly 151  
Db 3015 -----GCACACACA 3023  
QY 152 AlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrrGluPhe 171  
Db 3024 GCCTTTTCAGATGCCAGCTACAGCAGAACTACTTCTT-----TTT 3065  
QY 172 LeuLysSerLysArgGluGlu----- 178  
Db 3066 TTACGGAAAAGATGACCCAGCTTCCTAAAGGTCTCTTGGAAAGCTTGTCTGAAATGGCT 3125  
QY 179 -----PheGlnGluTyrrLeu-----GlnLysLeu 186  
Db 3126 GACGCTCTCTATATAGCTCTACCTGTTATTAGGAAACTTATCGCAACCAATCTATC 3185  
QY 187 LeuGlnHisProGluLeuSerAsnSerGln----- 196  
Db 3186 CTGAGTTCTCCAGACTTGTCCAGTTCAAGTCTATATTTAAGAAAACGGAAGTTCT 3245  
QY 196 ----- 196  
Db 3246 CTGGAGCCACGAAACTATCTCAAGACTTTCTTAAAAATATTCTCTGTGTGAACCTTGA 3305  
QY 197 -----LeuLeuAlaLap----- 200  
Db 3306 GTGAGGAGGAGAACACAGAAAGTTTGTCTAAAGAGTTGGCTTGTCTATTGAAGATGAACCT 3365  
QY 201 -----PheLeuSerProAsnGly---GlyGluThrGlnPheLeuAspLysIleLeu 216  
Db 3366 GTCTCTCTGATTCAGCGCCGACAGGGATGGGAAAACC-----TATGGCTATCTCTTA 3419  
QY 217 ProAspValAsnLeuGlyLys-----IleIleLysSerValProGlyLys----- 231  
Db 3420 CCGCTTTATCTCAATCCAAAGAGCGACAAATTTGTTCTTAGTGTCCGACAAAGATCTT 3479  
QY 232 -----LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIle 248  
Db 3480 CAAAATCAAATCATGGAAGAAGGTAAACGGCTCAAGGAAGTGTCCATACAGATATT 3539  
QY 249 AsnSerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerPro 268  
Db 3540 CATAGCTTAAAGGGACCA----- 3557  
QY 269 ThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAla 288  
Db 3558 -----CAAAATATCTGAGTTG-----GATGCTTTTATCATCTCTTGGAG----- 3599  
QY 289 GluAsnThrGluArgLysGlnAsnGlnAsnTyrrPheMetGluValMet----- 304  
Db 3600 GAAAATGATGAAAATCGCTTATTAGCGCTTTAAATGCAAGTCTGTGCTGCTTACT 3659  
QY 305 ---ThrValGluGlyValTyrrAspTyrrLeuMetTyrrValGlyArgValValPheGlnVal 323  
Db 3660 GAGACAGACAGACAGATTTGGATGAAATCGGGCAACTCTACCGTTACCAACATTTTCTA 3719  
QY 324 ProAspTrpLeuHisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGlu 343  
Db 3720 GCAGACCTTCGTATGATGGAAATTTATCATCCAGAGCTTATTT----- 3764  
QY 344 MetTyrrThrAspTyrrTyrrLeuGln-----CysLysLeuGluGln 356

```
Db 3765 GTGACGGAAGATTTTTGGAAACGTAGTCAAGAAAGGCGAGACTTGCAGCTT----- 3818
Qy 357 LeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaIlePheCys 376
Db 3819 TTAGTGAATATCAT-----GCCTATCTCTGATACCAAGCTTGAAGAT----- 3860
Qy 377 GluAsnThrGluProArgSerLeuGluAspGlnLysGlyAlaLysGlnThrPheGlu 396
Db 3861 -----AATCTGAAATTTGTCAGTGACCGT----- 3884
Qy 397 GluMetMetAsnTyrIleProAspLeuValLysCysIleGlyGluGluThrLysTyr 416
Db 3885 -----TTACTGATT-----ATTGATGAAGTCCAAG----- 3911
Qy 417 GluSerLeuArgLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436
Db 3912 -----ATTTTGTAGCTCTAGAAATCTGCTTCAAGAGACCTACGATATACAA----- 3959
Qy 437 TyrValLeuLeuAspIleValIleGlnGluPheProGluLeuAsnLysValGlnLys 456
Db 3960 ---TCTATTATCGATTAAATTGATAGGCTTTAGTAGGAGAGAAACAGGGTTCAACAA 4016
Qy 457 GluVal 458
Db 4017 CGGATA 4022

RESULT 12
US-09-635-872A-21
; Sequence 21, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 21
; LENGTH: 3621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-635-872A-21

Alignment Scores:
Pred. No.: 0.0225 Length: 3621
Score: 108.00 Matches: 104
Percent Similarity: 33.94% Conservative: 63
Best Local Similarity: 21.14% Mismatches: 153
Query Match: 4.43% Indels: 172
Dels: 4 Gaps: 29

US-09-744-313A-1 (1-465) x US-09-635-872A-21 (1-3621)

Qy 4 IleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSer 23
Db 1813 ATACATTATATCTACGTTTCAAAATACA-----AACGATGAAGATTC----- 1857
Qy 24 ArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43
Db 1858 -----AAATCCAGATACACAGTGGAGTGCACACAGGGAGGACCAAACTG 1905
Qy 44 ProAsnTyrGlyValAlaGluGlyCysAspPheIleGlu----- 57
Db 1906 ---AATTCTAGCCGCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACACTTCTTCCA 1962
Qy 58 GluGlyIleValValMetGluAspSerProValGluAlaValSerThrProAsnThr 77
Db 1963 GAGCGGGAATCTCTTTTGAGGAT-----ATTCCCAAGAA 1998
```

```
Qy 78 ProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPhePheGluAspPro 97
Db 1999 AAACGC-----TTCCCGAAGTTTAAAGACACATGATTTAAC 2034
Qy 98 SerSerGluArgLysGlyLysGluArgIleProValPheCysIleAspValGluArg 117
Db 2035 TCAACAGAGAGAGCCAGAGAGTGAATAATTCCTCGTAAATATTTTCACTCCTTCCA 2094
Qy 118 AsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeu 137
Db 2095 AAAGACGCCAG-----TTG 2109
Qy 138 GluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGln 157
Db 2110 AGTCTCAATACCTTGATTCCAACTGGA-----CATGGA-----GACATCACT 2154
Qy 158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGlu 177
Db 2155 TTG-----AAAGGATACAATTTGTCCAAAGTCA----- 2181
Qy 178 GluPheGlnGluTyrLeuGlnLysLeuLeu-----GlnHisProGluLeuSerAsn 194
Db 2182 -----GCCTTGTCTGAGATCATTTCTGATGAACCTCACAGCATGCTAAATAAATAA 2232
Qy 195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209
Db 2233 CAAGCTATATAACAGATGAACAAATGACAGTTGGTGGCTCCACAG-----GAAAA 2286
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLys----- 226
Db 2287 CAGGTTTCATAAAGCATCTTGCCTAAC-----AGCTTAGGAGTGTCTGAAGATTCACAGG 2343
Qy 227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis----- 239
Db 2344 TTGACTTTCTCTGCGATGAGTAAAGTGAATGTCATGACCAGGCTCAGAAATCCACCC 2403
Qy 240 -----LeuGluProPheIleMetAsnPheIleAsn 249
Db 2404 CTGGACTTGGAGACACACAGCAAGATTAGAGTGAACACTCACCCAAAAAACCATAGGC 2463
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
Db 2464 GGA---AATGTGACAAAGAAAGCC-----CCATCTCTGATGTTTCCACTGGAAAGC 2514
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
Db 2515 CAGATGACAAAAAGAAAGAAATACAGGGGAAAGAAAGAGACAGTAGAATGAGGAA 2574
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309
Db 2575 AATGCTGAAAA----- 2586
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis 329
Db 2587 -----CACATAGGC----- 2595
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349
Db 2596 -----GTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATACACAGATAGTTAC 2646
Qy 350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361
Db 2647 TTGGGCTTTTGGCATGGAGAAAAAAGTATTTCCTAGATCTTTCGACGAAGAGAG 2706
Qy 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381
Db 2707 TCATTGAAGACACAATGGCATCTTCACTGATAGC-----AAGAACTACT---GGG 2754
Qy 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetAsnTyr 401
Db 2755 AGGCACTAAAGAT-----ACATTTGCAGATTCCTCAGATAT 2793
Qy 402 IleProAspLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArgLeu 421
```

Db 2794 GTAAATAAAATCTTAAATAGCAAGTTTGA----- 2823  
Qy 422 LeuPheaspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyValLeuLeuasp 441  
Db 2824 ---TTACATCGCGAAGGTCCTGCT-----CACATGCCCTCACATGATGACCGG 2871  
Qy 442 IleValIleGlnGluLeu-----PheProGlu 450  
Db 2872 ATTGTATGCAAGAACTGCAAGATATGTTCCCTGAA 2907

## RESULT 13

US-09-636-077A-21

; Sequence 21, Application US/09636077A

; Patent No. 653785

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE

; FILE REFERENCE: 195612US0

; CURRENT APPLICATION NUMBER: US/09/636,077A

; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 21

; LENGTH: 3621

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-636-077A-21

## Alignment Scores:

Pred. No.:	0.0225	Length:	3621
Score:	108.00	Matches:	104
Percent Similarity:	33.94%	Conservative:	63
Best Local Similarity:	21.14%	Mismatches:	153
Query Match:	4.43%	Indels:	172
DB:	4	Gaps:	29

US-09-744-313A-1 (1-465) x US-09-636-077A-21 (1-3621)

Qy 4 IleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSer 23  
Db 1813 ATACATTTTAACTCAGCTTCAAAATACA-----AAGATGAAGGTTCT----- 1857  
Qy 24 ArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43  
Db 1858 -----AAATGCAGATAACAGTGGAGGTGGACACACAGGAGGACCAAACTG 1905  
Qy 44 ProAsnTyrGlyValAlaGluGlyGluAspAspPheIleGlu----- 57  
Db 1906 ---AATTCTACGGCCAGAGAGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCA 1962  
Qy 58 GluGlyIleValValMetGluAspAspSerProValGluAlaValSerThrProAsnThr 77  
Db 1963 GAGCGGAAATCCTTTTGGAGT-----ATTCCCAAGAA 1998  
Qy 78 ProArgAsnLeuAlaAlaTrpLysIleSerIlePheProTyrValAspPheGluAspPro 97  
Db 1999 AAACGC-----TTCCGGAAGTTTAAAGACATGATGTTAA 2034  
Qy 98 SerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArg 117  
Db 2035 TCAACAGGAGAGCCAGGAGAGGTTGAAATTTCCCTCGTAAATATTTTCACTCTCTTCCA 2094  
Qy 118 AsnAspArgArgAlaValAlaGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeu 137  
Db 2095 AAAGACCCCGAG-----TTG 2109  
Qy 138 GluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGln 157  
Db 2110 AGTCTCAATACCTTGGATTTGCAACTGGAA-----CATGGA-----GACATCACT 2154

Qy 158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGlu 177  
Db 2155 TTG-----AAAGGATACAAATTTGTCCAGTCA----- 2181  
Qy 178 GluPheGlnGluTyrLeuGlnLysLeuLeu-----GlnHisProGluLeuSerAsn 194  
Db 2182 -----GCCTTGCTGAGATCATTTCTGATGAACATCAGCATGCTAAATAAAAAAT 2232  
Qy 195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209  
Db 2233 CAAGCTATAATACAGATGAAACAAATGACAGTTTGTGGCTCCACAG-----GAAAAA 2286  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLys----- 226  
Db 2287 CAGGTTCAATAAAGCACTTGGCCAAAC---AGCTTAGGAGTGTCTGAAAGATTGCAGAGG 2343  
Qy 227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis----- 239  
Db 2344 TTGACTTTTCTCGCAGTGAGTGTAAAGAGTGAATGAGTGCATGACCCAGGTCAGATCCACCC 2403  
Qy 240 -----LeuGluProPheIleMetAsnPheIleAsn 249  
Db 2404 CTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAACCATAGGC 2463  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 2464 GGA---AATGTGACAAAGAAAGAGCCC-----CCATCTCTGATTGTTCCTCGGAAGC 2514  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 2515 CAGATGACAAAGAAAGAAATTCACAGGGGAAAGAAAGAGAACAGACAGTAGAATGGAGAA 2574  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 2575 AATGCTGAAAT----- 2586  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrpLeuHisHis 329  
Db 2587 -----CACATAGGC----- 2595  
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 2596 -----GTTACTGAGTGTACTTGGAAAGAAAGCTGCAGCATTACACAGATAGTAC 2646  
Qy 350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361  
Db 2647 TTGGCTTTTTCATCGGAGAAAAAGTATTTCTAGATCTTCTCGACGAAGAGAG 2706  
Qy 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381  
Db 2707 TCATTGAAGACACAAATTTGGCATACTTCCACTGATAGC-----AAGATACT---GGG 2754  
Qy 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401  
Db 2755 AGGCACTTAAAGAT-----ACATTTGCAGATTCCCTCAGATAT 2793  
Qy 402 IleProAspLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArgLeu 421  
Db 2794 GTAAATAAATCTTAAATAGCAAGTTTGA----- 2823  
Qy 422 LeuPheaspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441  
Db 2824 ---TTCACATCGCGAAGATCCCTGCT-----CACATGCTCATCATGATTGACCGG 2871  
Qy 442 IleValIleGlnGluLeu-----PheProGlu 450  
Db 2872 ATTGTATGCAAGAACTGCAAGATATGTTCCCTGAA 2907

## RESULT 14

US-09-635-872A-20

; Sequence 20, Application US/09635872A

; Patent No. 6534300

; GENERAL INFORMATION:

APPLICANT: CANFIELD, WILLIAM  
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

FILE REFERENCE: 195612US0  
CURRENT APPLICATION NUMBER: US/09/635,872A

CURRENT FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831

PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52

SOFTWARE: Patent in version 3.0

SEQ ID NO 20

LENGTH: 3783

TYPE: DNA

ORGANISM: Homo sapiens

US-09-635-872A-20

#### Alignment Scores:

Pred. No.:	0.0242	Length:	3783
Score:	108.00	Matches:	104
Percent Similarity:	33.94%	Conservative:	63
Best Local Similarity:	21.14%	Mismatches:	153
Query Match:	4.43%	Indels:	172
DB:	4	Gaps:	29

US-09-744-313A-1 (1-465) x US-09-635-872A-20 (1-3783)

```
QY 4 IleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGlySerPheGlyIleSer 23
DB 1837 ATACATTTAATCTCAGCTTTCAAAATACA-----AACGATGAAGAGTTC----- 1881
QY 24 ArgIleGlySerIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43
DB 1882 -----AAAATCGCATACACAGTGGAGTGACAGCAAGGGAGGAGCCAAACTG 1929
QY 44 ProAsnTyrGlyValAlaGluGlyGluAspPheIleGlu----- 57
DB 1930 ---AATTCACGGCCAGAGGGTACGAAATTTAGTTAGTCCCATACACTTCTCCA 1986
QY 58 GluGlyIleValMetGluAspSerProValGluAlaValSerThrProAsnThr 77
DB 1987 GAGCGGAAATCCCTTTTGAGAT-----ATTCCCAAGAA 2022
QY 78 ProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPhePheGluAspPro 97
DB 2023 AAACGC-----TTCCGGAAGTTTAAGACACATGATGTTAAC 2058
QY 98 SerSerGluArgLysGluLysGluArgIleProValPheCysIleAspValGluArg 117
DB 2059 TCACACAGGAGCCCGAGAGAGGTGAATTCCTGTAATATTTCACTCCTTCCA 2118
QY 118 AsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeu 137
DB 2119 AAAGAGCCCGAG-----TTG 2133
QY 138 GluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGln 157
DB 2134 AGTCTCAATACCTTGGATTTGCAACTGAA-----CATGGA-----GACATCACT 2178
QY 158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLysSerLysArgGlu 177
DB 2179 TTG-----AAAGATACAAATTTGTCCAAGTCA----- 2205
QY 178 GluPheGlnGluTyrLeuGlnLysLeu-----GlnHisProGluLeuSerAsn 194
DB 2206 -----GCCCTGCTGAGATCATTTCTGTAGTCACTACAGCATGTCTAAATAAATAAT 2256
QY 195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209
DB 2257 CAAGCTAATAACACAGATGAACAAATGACAGTTTGGTGGTCCACAG-----GAAAAA 2310
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLys----- 226
DB 2311 CAGGTTTCATAAAGCATCTTCCCAAC-----AGCTTAGGAGTGTCTGAAGATTGCAGAGG 2367
```

```
QY 227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis----- 239
DB 2368 TTGACTTTTCTCGAGTCAGTGAATAAGTGAATCGTATGACCCAGGTGCAATCCACCC 2427
QY 240 -----LeuGluProPheIleMetAsnPheIleAsn 249
DB 2428 CTGGACTTGGAGACACACAGCAAGATTAGAGTGGAAACTCACCCAAAAAACCATAGGC 2487
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
DB 2488 GGA---AATGTGACAAAAGAAAAGCCC-----CCATCTCTGTGTTGTTCCTCGAAAGC 2538
QY 270 SerGluAsnAsnLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
DB 2539 CAGATGACAAAAGAAAAGAAAATCACAGGGGAAAAGAAAAGACAGTAGAATGAGAGAA 2598
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309
DB 2599 AATGCTGAAT----- 2610
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHis 329
DB 2611 -----CACATAGGC----- 2619
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349
DB 2620 -----GTTACTGAAGTCTTACTTGGAAAGAGTGCAGCATACACAGATAGTTAC 2670
QY 350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361
DB 2671 TTGGCTTTTGGCCATGGGAGAAAAGTATTTCTTAGATCTTCTGCACCAAGAGAG 2730
QY 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381
DB 2731 TCATTGAAGACACAAATGGCATACTTCACTCATAGC-----AAGAATACT---GGG 2778
QY 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401
DB 2779 AGGCAACTAAAAGAT-----ACATTTGCAGATTCCCTCAGATAT 2817
QY 402 IleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeu 421
DB 2818 GTAAATAAATTTCTAAATAGCAAGTTTGA----- 2847
QY 422 LeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441
DB 2848 ---TTCACATCGCGAAAGTCCCTGCT-----CACATGCCTCACATGATTGACCGG 2895
QY 442 IleValIleGlnGluLeu-----PheProGlu 450
DB 2896 ATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAA 2931
```

#### RESULT 15

US-09-636-077A-20

Sequence 20, Application US/09636077A

Patent No. 6537785

GENERAL INFORMATION:

APPLICANT: CANFIELD, WILLIAM

TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE

FILE REFERENCE: 195612US0

CURRENT APPLICATION NUMBER: US/09/636,077A

CURRENT FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831

PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52

SOFTWARE: Patent in version 3.0

SEQ ID NO 20

LENGTH: 3783

TYPE: DNA

ORGANISM: Homo sapiens

US-09-636-077A-20

Alignment Scores:

Pred. No.: 0.0242 Length: 3783  
Score: 108.00 Matches: 104  
Percent Similarity: 33.94% Conservative: 63  
Best Local Similarity: 21.14% Mismatches: 153  
Query Match: 4.43% Indels: 172  
DB: 4 Gaps: 29

US-09-744-313A-1 (1-465) x US-09-636-077A-20 (1-3783)

QY 4 ILeHisPheCysLeuLeuPheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSer 23  
DB 1837 ATACATTTTAACTCAGCTTCAAAATACA-----AACGATGAAGAGTTC-----1881

QY 24 ArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43  
DB 1882 -----AAATGCGAGTAAACAGTGGAGGTGACACAGGGAGGAGCAAAACTG 1929

QY 44 ProAsnTyrGlyValAlaGluGlyGluAspPheIleGlu-----57  
DB 1930 ---AATTCACGGCCAGAGGGTTACGAAATTTAGTTAGTCCCAATACACTTCTTCCA 1986

QY 58 GluGlyIleValValMetGluAspSerProValGluAlaValSerThrProAsnThr 77  
DB 1987 GAGCGGAAATCCTTTTTCAGGAT-----ATTCCCAAGAA 2022

QY 78 ProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPheGluAspPro 97  
DB 2023 AAACGC-----TTCCCGAAGTTTAAAGACATGATGTTAAC 2058

QY 98 SerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArg 117  
DB 2059 TCACAGGAGAGCCAGGAGAGGTGAAATTCCTCGTAAATATTTCACTCCTTCCA 2118

QY 118 AsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeu 137  
DB 2119 AAAGAGCCCGAG-----TTG 2133

QY 138 GluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAsnAlaGln 157  
DB 2134 AGTCTCAATACCTTGGATTGCAACTGGAA-----CATGGA-----GACATCACT 2178

QY 158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGlu 177  
DB 2179 TTG-----AAAGNATCAATTTGCTCCAGTCA-----2205

QY 178 GluPheGlnGluTyrLeuGlnLysLeuLeu-----GlnHisProGluLeuSerAsn 194  
DB 2206 -----GCCTTGCTGAGATCATTTCTGATGAACCTCAGACATGCTAAATAAAAT 2256

QY 195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209  
DB 2257 CAAGCTATATAACAGATGAACAACAAATGACAGTTTGGTGGCTCCACAG-----GAAAAA 2310

QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLys-----226  
DB 2311 CAGGTTCAAAAAGCATCTTGGCAAAAC---AGCTTAGAGTGTCTCAAAAGATTGACAGG 2367

QY 227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis-----239  
DB 2368 TTGACTTTTCTGAGTGAGTGTAAGAAGTGAATGGTCAAGCAGGCTCAGATCCACCC 2427

QY 240 -----LeuGluProPheIleMetAsnPheIleAsn 249  
DB 2428 CTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAAACCATAGGC 2487

QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
DB 2488 GGA---AATGTGCAAAAGAAAGAAAGCC-----CCATCTCTGATTTGCTCCACTGGAAAGC 2538

QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
DB 2539 CAGATGCAAAAGAAAGAAAGAAATCAACAGGGAAGAAAGAGAACAGATAGATGGAGGAA 2598

QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
DB 2599 AATGCTGAAAAT-----2610

QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisHis 329  
DB 2611 -----CACATAGGC-----2619

QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 2620 -----GTTACTGAGTGTACTTGGAGAAAGCTGACGATTACACAGTAGTTAC 2670

QY 350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361  
DB 2671 TTGGCGCTTTTGGCATGGGAGAAAAAAGTATTTCTAGATCTTCTCGACGAGAGAG 2730

QY 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381  
DB 2731 TCATTGAAGACACAAATTGGCATACTTCACTGATAGC-----AAGATACT---GGG 2778

QY 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401  
DB 2779 AGGCAACTAAAGAT-----ACATTTGCAGATTCCTCCAGATAT 2817

QY 402 IleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeu 421  
DB 2818 GTAATATAAATCTTAATACCAAGTTTGG-----2847

QY 422 LeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441  
DB 2848 ---TTCAATCGCGGAAAGTCCCTGCT-----CACATGCTCAGATGATTGACCGG 2895

QY 442 IleValIleGlnGluLeu-----PheProGlu 450  
DB 2896 ATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAA 2931

Search completed: January 31, 2004, 05:18:33

Job time : 523 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 06:31:47 ; Search time 86 Seconds  
(without alignments)  
2386.548 Million cell updates/sec

Title: US-09-744-313A-1  
Perfect score: 465  
Sequence: 1 MYLIRCLIFRNTQKGSF.....ELFPELNKVKQVTSWMM 465

Scoring table: OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DSV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113613\_4617/app\_query.fasta\_1.647  
-DB=Issued Patents NA -QPMF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMF=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 -CGN\_1=56 -runat\_30012004\_113613\_4617 -NCFPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/2/ina/5A COMB.seq:  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq:  
3: /cgn2\_6/prodata/2/ina/6A COMB.seq:  
4: /cgn2\_6/prodata/2/ina/6B COMB.seq:  
5: /cgn2\_6/prodata/2/ina/PTUS COMB.seq:  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	455	98.1	1716	4	US-09-620-312D-922
2	307	66.0	1551	4	US-09-620-312D-923
3	8	1.7	85	2	US-08-332-766A-44
4	8	1.7	740	4	US-09-342-681C-99
5	8	1.7	966	4	US-09-252-991A-9495
6	8	1.7	1239	4	US-09-252-991A-9666
7	8	1.7	1842	4	US-09-205-258-160
8	8	1.7	3043	3	US-09-008-271A-14
9	8	1.7	152331	3	US-09-128-155-16
10	8	1.7	176373	3	US-09-128-155-17
11	7	1.5	161	1	US-08-110-786A-4
12	7	1.5	161	1	US-08-110-786A-6

c 13	7	1.5	223	1	US-08-435-684A-43	Sequence 43, Appl
c 14	7	1.5	223	2	US-08-934-877A-43	Sequence 43, Appl
c 15	7	1.5	223	3	US-08-871-678C-43	Sequence 43, Appl
c 16	7	1.5	276	4	US-09-252-991A-991	Sequence 991, Appl
c 17	7	1.5	280	4	US-09-313-294A-998	Sequence 998, Appl
c 18	7	1.5	285	4	US-09-107-532A-3549	Sequence 3549, Appl
c 19	7	1.5	288	4	US-09-313-294A-3533	Sequence 3533, Appl
c 20	7	1.5	288	4	US-09-313-294A-5146	Sequence 5146, Appl
c 21	7	1.5	321	1	US-08-175-388-2	Sequence 2, Appl
c 22	7	1.5	321	2	US-08-779-620-2	Sequence 2, Appl
c 23	7	1.5	321	2	US-08-818-726-2	Sequence 2, Appl
c 24	7	1.5	321	3	US-09-018-584A-20	Sequence 20, Appl
c 25	7	1.5	365	3	US-09-326-039-1	Sequence 1, Appl
c 26	7	1.5	370	2	US-08-332-766A-8	Sequence 8, Appl
c 27	7	1.5	371	3	US-09-326-039-11	Sequence 11, Appl
c 28	7	1.5	421	2	US-08-332-766A-25	Sequence 25, Appl
c 29	7	1.5	475	4	US-09-221-017B-561	Sequence 561, Appl
c 30	7	1.5	597	2	US-08-332-766A-19	Sequence 19, Appl
c 31	7	1.5	627	3	US-09-328-111-194	Sequence 194, Appl
c 32	7	1.5	657	4	US-09-252-991A-13015	Sequence 13015, A
c 33	7	1.5	661	2	US-08-529-978B-37	Sequence 37, Appl
c 34	7	1.5	666	4	US-09-252-991A-11270	Sequence 11270, A
c 35	7	1.5	702	4	US-09-252-991A-6791	Sequence 6791, Appl
c 36	7	1.5	710	3	US-08-998-416-982	Sequence 982, Appl
c 37	7	1.5	725	3	US-08-998-416-984	Sequence 984, Appl
c 38	7	1.5	765	4	US-09-252-991A-6948	Sequence 6948, Appl
c 39	7	1.5	780	4	US-09-252-991A-6722	Sequence 6722, Appl
c 40	7	1.5	784	4	US-09-205-258-101	Sequence 101, Appl
c 41	7	1.5	797	3	US-08-981-803-13	Sequence 13, Appl
c 42	7	1.5	797	3	US-08-981-803-27	Sequence 27, Appl
c 43	7	1.5	797	3	US-08-983-440-13	Sequence 13, Appl
c 44	7	1.5	797	3	US-08-983-440-27	Sequence 27, Appl
c 45	7	1.5	797	4	US-09-367-895-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1  
US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yundong  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (71)..(1543)  
US-09-620-312D-922

Alignment Scores:  
Pred. No.: 0  
Score: 456.00  
Length: 1716  
Matches: 456  
Percent Similarity: 100.00%  
Best/Local Similarity: 100.00%  
Query Match: 98.06%  
DB: 4  
Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
Db 173 TTTTCGGAACACACAGAAAGGGAGACATATTGGAAATCAGCAGAAATAGGTAGCAAAAT 232  
QY 30 LysGlyValPheLysSerThrMetGluGlyAlaMetLeuProAsnThrGlyValAla 49  
Db 233 AAAGAGATATCCAAAGTACCAATAGGAGGAGCTATGTTGCCCTAAATATAGGTGTAGCT 292  
QY 50 GluGlyGluAspPheIleGluGlyIleValValMetGluAspSerProVal 69  
Db 293 GAGGTGAGATGATTTTATTAAGAGAGGTATTTGTTAATGGAAGATGATTTCCAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaIleLysIleSerIlePro 89  
Db 353 GAGGCTGTGAGCACACCTAAATCTCCCGAAACCTTGCTGCATGCAAAATATAGCATTTCCA 412  
QY 90 TyrValAspPheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro 109  
Db 413 TATGTAGACTTTTTTGGAGATCCCTCTCTGAAGAGAGGAGGAAAGAAAGAAATTCCT 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 473 GGTGTTTGTATGATTTGAAGAATAATGATAGAGAGCAGGTGGACAGACCTTGAAAT 532  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 533 TGGTCTGCTCTAGAGATATCTTGAATCTATGTACTTGAATCAAACTAAACAGAAATTT 592  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 593 CATGTGTGATTTCTCTGATGCCAGCTTCTCTTAAGAGGATCAATGGCCCCCAAAATAT 652  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
Db 653 GAATTTCTTAAAGTCAAGAGGAGGAGGATTCGAAGAATATCTACAGAACTTTCTGCAGAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 713 CCAGAACTGAGTAATAGTCACTTCTGCAGACTTTCTTCCCTTAATGGTGGGGAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229  
Db 773 CAATTTCTTGAAGATACCTACAGATGTAATCTTTGGGAAAATTTATAAAATCTGTTCCT 832  
QY 230 GlyLysLeuMetLysGlyGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
Db 833 GGAAGAACTAATGAAGAGAGAGGTCAGCATTTGGAACCTTTTATCATGAAATTTCAATAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 893 TTTTGTGAGTCTCCAAAGCTTAACCAAGTAGACCAAGCACTGACCATTTCTCAGCCCTACT 952  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
Db 953 TCAGAAAACACACAGAGCTTTTCAATGATCTGTTTAAATATATGCAACCGTGTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309

Db 1013 AATACAGAGAGAAAGCAAAATCAGAAATATATTTATGGAGGTGATGACTGTAGAGAGTC 1072  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis 329  
Db 1073 TATGATTACCTCATGTATGTAGGAGGGTAGTTTCCAGGTTCCCTGACTGGCTTCATCAT 1132  
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 1133 CTCCTTAATGGGAATCGAATCTCTTTAAAACACCCCTGGAATGTATGATTACTAT 1192  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
Db 1193 CTTCAAGTAAACTAGAACAGCTATTTTCCAGGAGCACCGTTTGGTCTCACTCATAACACTT 1252  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
Db 1253 CTCAGAGATGCTATATTTCTGTGAAAACACCTGAACCTCGCTCTCTCAAGATTAAGCAAAA 1312  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
Db 1313 GGAGCAAAACACACTTTTGAAGAAATGATGAATATCATCCAGATCTGTAGTCAAGTGT 1372  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
Db 1373 ATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTGTGATGGCTTACAGCAACA 1432  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
Db 1433 GTACTCAACAGCAGCTGACTTATGTTTTATTGGACATTTGTGATACAGAACTGTTTCCA 1492  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
Db 1493 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATG 1540

RESULT 2  
US-09-620-312D-923  
Sequence 923, Application US/09620312D  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyang  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunding  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
FILE REFERENCE: 784CIP28  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt.PL\_genes version 1.0  
SEQ ID NO 923  
LENGTH: 1551  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:

NAME/KEY: CDS  
LOCATION: (71)..(1378)  
US-09-620-312D-923

Alignment Scores:

Pred. No.: 3 15e-306 Length: 1551  
Score: 307.00 Matches: 307  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 56.02% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-923 (1-1551)

Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
Db 173 TTTCGGAACACACAGAAAGGGAGAAATCATTTGGTAATCAGCAGATAGTAGCAAAATT 232  
Qy 30 LysGlyValPheLysSerThrMetGluGlyAlaMetLeuProAsnThrGlyValala 49  
Db 233 AAAGGAGTATTCAAAAGTACCAATGGAGGGAGCTATGTTGCCCTAAATTATGTTAGCT 292  
Qy 50 GluGlyGluAspPheIleGluGlyIleValValMetGluAspSerProVal 69  
Db 293 GAAGGTGAGATGATTTTATGGAAGAGTATTTGTAATGGAAGATGATTCGAGTG 352  
Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
Db 353 GAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGCTGCATGGAAATAGCATTCACA 412  
Qy 90 TyrValaspPheGluAspProSerSerGluArgLysGluLysIleGluArgIlePro 109  
Db 413 TATGTAGACTTTTTCGAGTATCCTCTCTGAAAGGAGGAGAGAGAGAGAGAGAGAGAG 472  
Qy 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 473 GTGTTTGTATTCATGTTGAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532  
Qy 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 533 TGCTCTGTCTATGAGAGATCTTGTAATCTATGTAATCTGTAATCTGTAATCTGTAAT 592  
Qy 150 HisGlyValPheProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyr 169  
Db 593 CATGTGTATTCCTGATGCTCCAGCTTCTCTTCTAGAGAGATGATGCTCCGAGAGAT 652  
Qy 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
Db 653 GAATTCCTTAAGTCAAAGAGGAGAGAGATCTCAGAGATATCTCAGAGAGATCTCAGAG 712  
Qy 190 ProGluSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 713 CCGAAGCTGATGATATGATCAACTCTGCGAGAGCTTTCTTCCCTTAATGTTGGGAAACA 772  
Qy 210 GluPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
Db 773 CAATTCCTGATAGATGATCTACAGATGTAATCTTGGGAGAAATTAATAATCTGTTCT 832  
Qy 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheLeuMetAsnPhelLeu 249  
Db 833 GGAAGAACTAATGAAG 892  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 893 TCTGTGAGTCTCAAGAGCCTAAACCAAGTAGACAGAGAGAGAGAGAGAGAGAGAGAG 952  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 953 TCAGAAACACACAG 1012  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 1013 AATACAG 1072

Qy 310 TyrAspTyrLeuMetTyrVal 316  
Db 1073 TATGATTACCTGATGATGTA 1093

RESULT 3

US-08-332-766A-44  
; Sequence 44, Application US/08332766A  
; Patent No. 5843647  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREYS, Alec J.  
; APPLICANT: ARMOUR, John  
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS  
; NUMBER OF SEQUENCES: 125  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DAREY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/332,766A  
; FILING DATE: 01-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326052.9  
; FILING DATE: 21-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BIRD, Donald J. 25,323  
; REGISTRATION NUMBER:  
; REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 85 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-332-766A-44

Alignment Scores:  
Pred. No.: 6.59 Length: 85  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 2 Gaps: 0

US-09-744-313A-1 (1-465) x US-08-332-766A-44 (1-85)

Qy 100 GluArgLysGluLysLysGluArg 107  
Db 49 GAAAGAAAGAGAGAGAGAGAGAGAGAGAG 72

RESULT 4

US-09-342-681C-99  
; Sequence 99, Application US/09342681C  
; Patent No. 6355782  
; GENERAL INFORMATION:  
; APPLICANT: Zonana et al.  
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins  
; FILE REFERENCE: 52978  
; CURRENT APPLICATION NUMBER: US/09/342,681C

; CURRENT FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: 60/092,279  
; PRIOR FILING DATE: 1998-07-09  
; PRIOR APPLICATION NUMBER: 60/112,366  
; PRIOR FILING DATE: 1998-12-15  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 99  
; LENGTH: 740  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-342-681C-99

Alignment Scores:  
Pred. No.: 50.3 Length: 740  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-342-681C-99 (1-740)

Qy 235 GluysGlyGlnHisLeuGlnPro 242  
Db 515 GAAAGGGGCGACATCTGGAGCCT 538

## RESULT 5

US-09-252-991A-9495/c  
; Sequence 9495, Application US/09252991A  
; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9495  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-9495

Alignment Scores:  
Pred. No.: 64.6 Length: 966  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-252-991A-9495 (1-966)

Qy 422 LeupheAspGlyLeuGlnPro 429  
Db 351 CTCTTCGACGGTTTCGAGCACA 328

## RESULT 6

US-09-252-991A-9666  
; Sequence 9666, Application US/09252991A  
; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9666  
; LENGTH: 1239  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9666

Alignment Scores:  
Pred. No.: 81.5 Length: 1239  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-252-991A-9666 (1-1239)

Qy 422 LeupheAspGlyLeuGlnPro 429  
Db 661 CTCTTCGACGGTTTCGAGCACA 684

## RESULT 7

US-09-205-258-160  
; Sequence 160, Application US/09205258  
; Patent No. 6525174

## GENERAL INFORMATION:

; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/09/205,258  
; CURRENT FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892

EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 160  
LENGTH: 1842  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (19)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (62)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1793)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1834)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-205-258-160  
Alignment Scores:  
Pred. No.: 118  
Score: 8.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.72%  
Length: 1842  
Matches: 8  
Conservative: 0  
Mismatch: 0  
Indels: 0

DB: 4 Gaps: 0  
US-09-744-313A-1 (1-465) x US-09-205-258-160 (1-1842)  
QY 98 SerSerGluArgLysGluLysLys 105  
Db 161 AGCAGTGAAGAAAGAAAGAAA 184  
RESULT 8  
US-09-008-271A-14  
Sequence 14, Application US/09008271A  
Patent No. 6203979  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
Hillman, Jennifer L.  
Yue, Henry  
Guegler, Karl J.  
Corley, Neil C.  
Tang, Tom Y.  
Shah, Purvi  
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Fast-Seq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/008,271A  
FILING DATE: 16-Jan-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mohan-Peterson, Sheela  
REGISTRATION NUMBER: 41,201  
REFERENCE/DOCKET NUMBER: PF-0458 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3043 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TLYMNOT02  
CLONE: 447484  
SEQUENCE DESCRIPTION: SEQ ID NO: 14 :  
US-09-008-271A-14  
Alignment Scores:  
Pred. No.: 190  
Score: 8.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.72%  
DB: 3  
Length: 3043  
Matches: 8  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0  
US-09-744-313A-1 (1-465) x US-09-008-271A-14 (1-3043)  
QY 98 SerSerGluArgLysGluLysLys 105  
Db 291 AGCAGTGAAGAAAGAAAGAAA 314

```
RESULT 9
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Alignment Scores:
Pred. No.: 7,486+03 Length: 152331
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.72% Indels: 0
DB: 3 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-128-155-16 (1-152331)

QY 100 GluArgLysGluLysLysGluArg 107
Db 128560 GAAAGAAAGAAAGAAAGAAAGAA 128537

RESULT 10
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Alignment Scores:
Pred. No.: 8,586+03 Length: 176373
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.72% Indels: 0
DB: 3 Gaps: 0

Query Match: 1.72% Indels: 0
DB: 3 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-128-155-17 (1-176373)

QY 100 GluArgLysGluLysLysGluArg 107
Db 144583 GAAAGAAAGAAAGAAAGAAAGAA 144560

RESULT 11
US-08-110-786A-4
; Sequence 4, Application US/08110786A
; Patent No. 5443966
; GENERAL INFORMATION:
; APPLICANT: FAIRWEATHER, Neil Fraser
; APPLICANT: MAKOFF, Andrew Joseph
; TITLE OF INVENTION: Expression of tetanus toxin fragment C
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 5443966th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,786A
; FILING DATE: 23-AUG-1993 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/777,337
; FILING DATE: 29-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00943
; FILING DATE: 20-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8914122.0
; FILING DATE: 20 June 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J. Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..161
US-08-110-786A-4

Alignment Scores:
Pred. No.: 128 Length: 161
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.51% Indels: 0
DB: 1 Gaps: 0
```

US-09-744-313A-1 (1-465) x US-08-110-786A-4 (1-161)

Qy 264 ThrilleuserProThrsr 270  
Db 118 ACGATATTCTCCGACATCT 138

## RESULT 12

US-08-110-786A-6/c  
Sequence 6, Application US/08110786A  
Patent No. 5443966  
GENERAL INFORMATION:  
APPLICANT: FAIRWEATHER, Neil Fraser  
APPLICANT: MAKOFF, Andrew Joseph  
TITLE OF INVENTION: Expression of tetanus toxin fragment C  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhye P.C.  
STREET: 1100 No. 5443966th Glebe Road  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,786A  
FILING DATE: 23-AUG-1993 1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/777,337  
FILING DATE: 29-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB90/00943  
FILING DATE: 20-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8914122.0  
FILING DATE: 20 June 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary J. Wilson  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 117-134  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NLXN UR  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 161 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
ORIGINAL SOURCE:  
ORGANISM: Clostridium tetani  
US-08-110-786A-6

Alignment Scores:  
Pred. No.: 128  
Score: 7.00  
Length: 161  
Matches: 7  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Conservative: 0  
Query Match: 100.00%  
Mismatches: 0  
Indels: 0  
Gaps: 0  
DB: 1

US-09-744-313A-1 (1-465) x US-08-110-786A-6 (1-161)

Qy 264 ThrilleuserProThrsr 270  
Db 48 ACGATATTCTCCGACATCT 28

## RESULT 13

US-08-435-684A-43/c  
Sequence 43, Application US/08435684A  
Patent No. 5707802  
GENERAL INFORMATION:  
APPLICANT: Sandhu, Gurpreet S.  
APPLICANT: Kline, Bruce C.  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection  
TITLE OF INVENTION: and Identification of Fungi  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba Corning Diagnostics Corp.  
STREET: 63 No. 5707802th Street  
CITY: Medfield  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02052  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS 6.2  
SOFTWARE: Word 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,684A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
NAME: Morgenstern, Arthur S.  
REGISTRATION NUMBER: 28,244  
REFERENCE/DOCKET NUMBER: CCD-180  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508 359-3836  
TELEFAX: 508 359-3885  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Candida glabrata specific region of 28S gene.  
HYPOTHETICAL: No  
ANY-SENSE: No  
US-08-435-684A-43

Alignment Scores:  
Pred. No.: 174  
Score: 7.00  
Length: 223  
Matches: 7  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Conservative: 0  
Query Match: 100.00%  
Mismatches: 0  
Indels: 0  
Gaps: 0  
DB: 1

US-09-744-313A-1 (1-465) x US-08-435-684A-43 (1-223)

Qy 250 SerCysGluSerProLysPro 256  
Db 57 AGCTGGAGACTCCCAAGCCC 37

## RESULT 14

US-08-934-877A-43/c  
Sequence 43, Application US/08934877A  
Patent No. 5958693  
GENERAL INFORMATION:  
APPLICANT: SANDHU, Gurpreet S.  
APPLICANT: KLINE, Bruce C.  
TITLE OF INVENTION: DNA ISOLATION METHOD  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.



COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,877A  
FILING DATE: 22-SEP-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,684  
FILING DATE: 05-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Bern D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 080394/0108  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Candida glabrata  
US-08-934-877A-43

Alignment Scores:  
Pred. No.: 174 Length: 223  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.51% Indels: 0  
DB: 2 Gaps: 0

US-09-744-313A-1 (1-465) x US-08-934-877A-43 (1-223)  
QY 250 SerCyeGluSerProlyPro 256  
Db 57 AGCTGCGAGAGTCCCAAGCCC 37

RESULT 15  
US-08-871-678C-43/c  
Sequence 43, Application US/08871678C  
Patent No. 6180339  
GENERAL INFORMATION:  
APPLICANT: Sandhu, Gurpreet S, Kline, Bruce C  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection and Identification  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Chiron Diagnostics Corporation  
STREET: 63 No. 6180339th Street  
CITY: Medfield  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02052  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS 6.2  
SOFTWARE: Word 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,678C  
FILING DATE: 06-JUNE-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/373,127  
FILING DATE: 13-JANUARY-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/435,684  
FILING DATE: 05-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Morgenstern, Arthur S.  
REGISTRATION NUMBER: 28,244  
REFERENCE/DOCKET NUMBER: CCD-180CIP11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508 359-3836  
TELEFAX: 508 359-3885  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Candida glabrata specific region of 28S gene.  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-871-678C-43

Alignment Scores:  
Pred. No.: 174 Length: 223  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.51% Indels: 0  
DB: 3 Gaps: 0

US-09-744-313A-1 (1-465) x US-08-871-678C-43 (1-223)  
QY 250 SerCyeGluSerProlyPro 256  
Db 57 AGCTGCGAGAGTCCCAAGCCC 37

Search completed: January 31, 2004, 09:06:06  
Job time: 177 secs



XX 07-DEC-2000.  
 XX 26-MAY-2000; 2000WO-US14831.  
 XX 27-MAY-1999; 99US-0136740.  
 XX 16-JUN-1999; 99US-0139566.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Yue H, Tang YT, Azimzai Y;  
 XX WPI; 2001-041141/05.  
 XX Novel Sorting Nexin polypeptides and polynucleotides useful for  
 PT diagnosing and treating disorders associated with their expression e.g.  
 PT autoimmune disorders, smooth muscle cell disorders and cell  
 PT proliferative disorders  
 XX Claim 5; Page 82-83; 84pp; English.  
 XX The present invention relates to human sorting nexin (SNEKN).  
 CC Compositions containing SNEKN or agonists of SNEKN are useful for  
 CC treating a disease or condition associated with decreased expression of  
 CC functional SNEKN and compositions containing antagonists of SNEKN are  
 CC useful for treating a disease or condition associated with  
 CC overexpression of functional SNEKN. These can be immune  
 CC disorders for example inflammation, asthma, allergy, and AIDS,  
 CC neurological disorders, gastrointestinal disorders, smooth muscle cell  
 CC disorders, cancers and others. The SNEKN may also be used for somatic  
 CC or germline gene therapy.  
 XX Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;  
 SQ

Alignment Scores:  
 Pred. No.: 0 Length: 1992  
 Score: 465.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAC86397 (1-1992)

Qy 1 MetTyrLeuLeuHisPheCysLeuLeuPheArgAsnThrGlnLysArgGlyGluSerPhe 20  
 Db 116 ATGACTTGATACATTTTGTAAATATTCAGGAACACACAGAAAAGGGGAGAAATCATTT 175  
 Qy 21 GlyTleSerArgIleGlySerLysLleLysGlyValPheLysSerThrThrMetGluGly 40  
 Db 176 GGAATCAGCAGATAGGTAGCMAAAATTAAGGAGATATTCAAAAGTACCAATGGAGGGA 235  
 Qy 41 AlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGluGlyIle 60  
 Db 236 GCTATGTTGCTAATTATGTTAGTACCTGAAGGTGAAGATGATTTATTTAGAGAGATTT 295  
 Qy 61 ValValMetGluAspAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80  
 Db 296 GTTGTATGAGAGATGATTCAGTGGAGGCTGTGAGCACCTAATACTCCCGAAG 355  
 Qy 81 LeuAlaLalTrpLysLleSerIleProTyrValAspPheGluAspProSerSerGlu 100  
 Db 356 CTTGCTGATGCAAAATAGCATTCATATGATAGACTTTTTTGGAGTACCTCTCTGAA 415  
 Qy 101 ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120  
 Db 416 AGAAGAGAGAAAAGAAAGAAATTCCTGTGTTTGTATGATGTTGAAGAAATATAGTA 475  
 Qy 121 ArgAlaValGlyHisGluProGluHisTrpSerValTyrArgGlyTyrLeuGluPheTyr 140  
 Db 476 AGAGCAGTTGGACAGCAGCTGACATTTGCTCTCTATAGAGATATCTTGATTTAT 535  
 Qy 141 ValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSer 160

Db 536 GTACTTGAATCAAACTAAACAGAAATTCATGTCATTTCTGATGCCAGCTTCCTCT 595  
 Qy 161 LysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGln 180  
 Db 596 AAGAGGATCATTCGCCCAAAATTAATGAATCTTAAAGTCAAGAGGGAAGAGTTCCAA 655  
 Qy 181 GluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
 Db 656 GAATACTACAGAAACTTCTGCAGCATCCAGAACTGAGTAATAGTCAACTCTCTGGCAGAC 715  
 Qy 201 PheLeuSerProAsnGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220  
 Db 716 TTTCTTTCCCTAATGTCGGGAAACACAAATTTCTTGATAAGATATACAGATGTAAT 775  
 Qy 221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlnHisLeu 240  
 Db 776 CTTGGGAAATATAAATCTGTTCTTGGAAACTAATGAAGAGAAAGGTGAGCTTTG 835  
 Qy 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
 Db 836 GAACCTTTTATCATGAATTTTCAATTAATCTTGTGAGTCTCCAAAGCTTAACCAAGTAGA 895  
 Qy 261 ProGluLeuThrIleLeuSerProThrSerGluAsnLysLysLeuPheAsnAspLeu 280  
 Db 896 CCAGAACTGACCATTTCTCAGCCCTACTTCAGAAAACACAAAGAGCTTTTCAATGATCTG 955  
 Qy 281 PheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPhe 300  
 Db 956 TTTAAAAATATGCAAAACCGTGTGAAATACAGAGAGAAAGCAAAATCAAGATTTATTT 1015  
 Qy 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
 Db 1016 ATGGAGGTGATGACTGTAGAGGAGTCTATGATTTACTGTATGTAGGACGGTAGTT 1075  
 Qy 321 PheGlnValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsn 340  
 Db 1076 TTCCAGGTTCCTGACTGGCTTCATCATCTCTTAATGGAACTCGAATCTCTTTAAAAAC 1135  
 Qy 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGlnLeuPheGlnGlu 360  
 Db 1136 ACCCTGGAAATGTATACTGATTTACTTCTCAGTGTAACTAGAACAGCTATTTCAGAG 1195  
 Qy 361 HisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380  
 Db 1196 CACGTTTGGTCTCCTCATCACTCACTCTCAGAGATGCTATATCTGTGAACACTGAA 1255  
 Qy 381 ProArgSerLeuGlnAspLysGlnLysGlyValalylsGlnThrPheGluGluMetMetAsn 400  
 Db 1256 CTTGCTCTCTCCAGATAGCAAAAGAGGACAAAACAGACTTTTGAAGAAATCATGAAT 1315  
 Qy 401 TyrIleProAspLeuLeuValLysCysIleGlyGluIleThrLysTyrGluSerIleArg 420  
 Db 1316 TACATTCAGATCTGTTAGTCAAGTATTGTCAGAGAACCAAGATAGAAAGCATCAGA 1375  
 Qy 421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
 Db 1376 CTTCTGTTGATGCTTACAGCAACCACTACTCAACAGCAGCTGATTTATTTATTTG 1435  
 Qy 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
 Db 1436 GACATTTGATACAGAACTGTTTCCAGAGCTCAATTAAGGTACAAAGAGAGTTACCTCT 1495  
 Qy 461 ValThrSerTrpMet 465  
 Db 1496 GTGACATCTGGATG 1510

RESULT 2  
 AAI59029  
 ID AAI59029 standard; cDNA; 1716 BP.  
 XX  
 AC AAI59029;  
 XX

DT	22-OCT-2001	(first entry)	Db	173	TTTCGGACACACAGAAAGGGGAGAAATCATTTGGATCAGCAGATAGGTAGCAAAATT	232
XX	Human polynucleotide SEQ ID NO 1232.					
DE	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		Qy	30	LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla	49
XX	Peripheral nervous system; neuropathy; central nervous system; CNS;		Db	233	AAAGGAGTATTCAAAGGTACCAATGGAGGAGCTATGTTGCTTAATATGTTAGT	292
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		Qy	50	GluGlyGluAspPheIleGluGlyIleValMetGluAspSerProVal	69
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		Db	293	GAAGGTGAAGATGATTTTATTTGAAGAAGGTATGTTGAATGAAGATGATTCTCCAGTG	352
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		Qy	70	GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTyrIleSerIlePro	89
XX	leukaemia; ss.		Db	353	GAGGCTGTGAGCACACCTTAATATCTCCCGAAACCTTGTGCAATGGAATAGCATTC	412
OS	Homo sapiens.		Qy	90	TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluArgIlePro	109
XX	WO200153312-A1.		Db	413	TATGTAGACTTTTGTGAGGATCCCTCTCTGAAGAGGAGAGAGAGAGAGATTCCT	472
XX	26-JUL-2001.		Qy	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
PF	26-DEC-2000; 2000WO-US34263.		Db	473	GTGTTTGTATTCATGTTGAAAGAAATGATAGAGAGAGTTCGACAGAGCTGAACAT	532
XX	21-JAN-2000; 2000US-0488725.		Qy	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
PR	25-APR-2000; 2000US-0552317.		Db	533	TGCTCTGTCTATAGAAAGATATCTTGAATTCATGTTGAATCAAACTAACAGATTT	592
PR	09-JUL-2000; 2000US-0598042.		Qy	150	HisGlyValPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr	169
PR	19-JUL-2000; 2000US-0620312.		Db	593	CATGGTGCAATTCCTGATGCCAGCTTCCTTCTAGAGGATCATTTGGCCCCCAAAATAT	652
PR	03-AUG-2000; 2000US-0653450.		Qy	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis	189
PR	14-SEP-2000; 2000US-0662191.		Db	653	GAATTCCTAAAGTCAAAGAGGAGGAGGAGTTCCAAGAAATATCTACAGAACTTCGAGCAT	712
PR	19-OCT-2000; 2000US-0693036.		Qy	190	ProGluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
PR	29-NOV-2000; 2000US-0727344.		Db	713	CCAGAACTCAGTAATAGTCAACTTCCTGGCAGACTTCTTCTCCCTTAATGCTGGGGAACA	772
XX	(HYSE-) HYSEQ INC.		Qy	210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro	229
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		Db	773	CAATTTCTTGATGAAGATACCTACAGATGTAATCTTGGGAAATATATAAAATCTGTCTCT	832
PI	PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		Qy	230	GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
PI	Zhao QA, Zhou P, Goodrich R, Dmanac RT;		Db	833	GGAAACTATGAAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGATTTTATAT	892
XX	WPI; 2001-442253/47.		Qy	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
DR	P-PSDB; AAM39873.		Db	893	TCITTTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACT	952
XX	Novel nucleic acids and polypeptides, useful for treating disorders		Qy	270	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
PT	such as central nervous system injuries -		Db	953	TCAGAAAACAAACAGAGAGCTTTTCAATGATCTGTTTAAATAATGATCAACCGTGTGAA	1012
XX	Claim 1; SEQ ID NO 1232; 10078pp; English.		Qy	290	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and		Db	1013	AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGAGGTGATGCTGTAGAGAGGTC	1072
CC	the encoded polypeptides (AAM39642-AAM42213) with nootropic		Qy	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHisHis	329
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		Db	1073	TATGATTAATGAGTATGATGATGAGCGGTAGTTTCCAGGTCTCTGCTGCTTCATCAT	1132
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		Qy	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
CC	of the invention may be used to treat diseases of the peripheral nervous		Db	1133	CTCTTAATGGAACTCGAATCTCTCTTTAAATAACCCCTGGAATGATATGATGATTAAT	1192
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		Qy	350	LeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
CC	localised neuropathies and central nervous system diseases, such as		Db	1193	CTTCAGTGTAAACTAGAACAGCTATTTCAGGAGCACCCGTTGGTCTCATCATACACTT	1252
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		Qy	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		Db	1253	CTCAGAGATGCTATATCTCTGTGAAACACTGAACCTCGCTCTCTCCAGAGTAAGCAAAA	1312
CC	utilisation of the activities such as: Immune system suppression,					
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic					
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,					
CC	assays for receptor activity, arthritis and inflammation, leukaemias and					
CC	C.N.S disorders.					
CC	Note: The sequence data for this patent did not form part of the printed					
CC	specification.					
XX	Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;					
XX	Alignment Scores:					
XX	Pred. No.: 0	Length: 1716				
XX	Score: 456.00	Matches: 456				
XX	Percent Similarity: 100.00%	Conservative: 0				
XX	Best Local Similarity: 100.00%	Mismatches: 0				
XX	Query Match: 98.06%	Indels: 0				
XX	DB: 22	Gaps: 0				
XX	US-09-744-313A-1 (1-465) x AAI59029 (1-1716)					
Qy	10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle					

QY 390 GlyValAspGlnThrPheGluMetMetAsnTyrIleProAspLeuValLysCys 409  
D 1313 GAGCAAAACAGCTTTTGGAGAAATGATGATTCATCCAGATCTGTAGTCAAGTGT 1372  
QY 410 IleGlyGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
D 1373 ATTGGTGAAGAAACCAAGATGAAAGCATCAGACTTCTCTTTGATGGCTTACAGCAACA 1432  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuPhePro 449  
D 1433 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGATCAGGAATCTGTTTCCA 1492  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
D 1493 GAGCTCAATAGGTACAAAGAGTACTCTCTGTGACATCTTGGATG 1540

## RESULT 3

AAI60815/c  
ID AAI60815 standard; cDNA; 2176 BP.

XX AAI60815;

DT 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4804.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPT; 2001-442253/47.

XX P-PSDB; AAM41659.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4804; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;

Alignment Scores:  
Pred. No.: 0 Length: 2176  
Score: 456.00 Matches: 456  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.06% Indels: 0  
DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAI60815 (1-2176)

QY 10 PheArgSerThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29

D 2022 TTTCCGAACACACAGAAAGGGGAGAAATCAATTTGGAAATCAGCAGAAATAGGTAGCAAAAT 1963

QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49

D 1962 AAAGAGTATTCAAAAGTACCACAATGGAGGAGCTATGTTGCTTAATTATGCTGTAGCT 1903

QY 50 GluGlyGluAspPheIleGluGluGlyValValMetGluAspSerProVal 69

D 1902 GAAGTCAAGATGATTTTATTGAAGAGGTATTGTTGAATGGAAGATGATTCCTCCAGTG 1843

QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaLalaTrpLysIleSerIlePro 89

D 1842 GAGCTCTGAGCACACCTAATCTATCTCCCGAAACCTTCTCATGGAATAATAGCATTTCCA 1783

QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysValGluArgIlePro 109

D 1782 TATGTAGACTTTTGTAGGATCCCTCTCTGAAGGAGGAGAGAAAGAAAGAAATTCCT 1723

QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129

D 1722 GTGTTTGTATTGATTTGAAAGAAATGATAGAGAGCAGTGTGACACAGAGCTGAACAT 1663

QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149

D 1662 TGGTCTGTATAGAGATATCTTGAATTCCTATGATCTTGAATCAAAACTAACAGAAATTT 1603

QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerIleArgIleIleGlyProLysAsnTyr 169

D 1602 CATGGTGATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGCCCCCAAAATATAT 1543

QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189

D 1542 GAATCTTAAAGTCAAGAGAGGAGAGTTCAGAGATATCTACAGAACTCTTGACAT 1483

QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209

D 1482 CCAGAGCTGACTAATAGTCAACTTCTGCGAGACTTTCTTCCCTAATAGTGGGGAACA 1423

QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229

D 1422 CAATTTCTTGATAAGATACTACAGATGTAATCTTGGGAAATATATAAATCTGTTCCT 1363

QY 230 GlyLysLeuMetLysGluLysGlnHisLeuGluProPheIleMetAsnPheIleAsn 249

D 1362 GGAATACTAATGAAGAGAGAGGTCAGCATTTGAACTTTTATCATGAATTCATTAAT 1303

QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269

D 1302 TCTTGTGATCTCCAAAGCCCTAAACCAAGTAGACCACTGACCAATTCCTCAGCCCTACT 1243

270	QY	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgLaGlu	289
1242	Db	TCAGAAAAACAACAAGAGCTTTTCAATGATCTCTTTTAAAAAATAATGCAAAACCGTCTGAA	1183
290	QY	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGluVal	309
1182	Db	AATACAGAGAGAAGCAAAATCAGATTATTTTATGGAGTGATGACTGTAGAGGAGTC	1123
310	QY	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHis	329
1122	Db	TATGATTACCTGATGTATGTAGGACGGGTAGTTTTCAGGTTCTGACTGGCTTCATCAT	1063
330	QY	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
1062	Db	CTCTTAATGGGAACCTCGAATCTCTTTAAAAAACACCCCTGAAATGTATACTGATTACTAT	1003
350	QY	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
1002	Db	CTTCAGTGTAACATAGAACAGCTATTTCAGAGACCGGTTTGCTCTCACTCATACACTT	943
370	QY	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
942	Db	CTCAGAGATGCTATATCTGTGAAAAACACTGAACCTCGCTCTCCAGATAAGCAAAA	883
390	QY	GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys	409
882	Db	CGAGCAAAACAGACTTTTGAAGAAATGATGAATTACATTCAGATCTGTAGTCAAGTGT	823
410	QY	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro	429
822	Db	ATTGTTGAAGAAACCAAGTATGAAGAGCATCAGACTTCTGTTTGATGGCTTACAGCAACA	763
430	QY	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449
762	Db	GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGTGATACAGGAAGTGTTC	703
450	QY	GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet	465
702	Db	GAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATG	655
RESULT 4			
AAI60816/c			
ID	AAI60816 standard; cDNA; 2176 BP.		
XX	AAI60816;		
AC	(first entry)		
DT	Human polynucleotide SEQ ID NO 4805.		
XX	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Sly-brager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia; ss.		
XX	Homo sapiens.		
XX	WO200153312-A1.		
PN	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
XX	21-JAN-2000; 2000US-0468725.		
FR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		

1602 CATGGTGCATTTCTTGATGCCAGCTCTCTCTAGAGATCATTTGGCCCCAAAATAT 1543  
170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
1542 GAATTCCTTAAGTCAAGAGGGAAGAGTTCCAGAAATATCTACAGAACTTCTGCAGCAT 1483  
190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
1482 CCAGAACTCAGTAATAGTCAACTCTGGCAGACTTCTTCCCTTAATGGTGGGAAACA 1423  
210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValPro 229  
1422 CAATTTCTTGATAGATATCTACAGATGTAATCTTGGGAAATATTAATCTGTTCT 1363  
230 GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
1362 GGAATACTAATGAAGAGAAAGTCCAGCATTTGGAACTTTTATCATGAATTCATTAAT 1303  
250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
1302 TCTTGTGAGTCTCCAAAGGCTAAACCAAGTAGACCAAGAACTGACCATTCTCGACCTACT 1243  
270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
1242 TCAGAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 1183  
290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
1182 AATACAGAGAAAGCAAAATCAGATTTATTTATGGAGTGATGATCTAGAGAGTCT 1123  
310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHisHis 329  
1122 TATGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063  
330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
1062 CTCTTAATGGAACTCGAATCTCTTTAAATAACACCTCGAATGATGATGATGATGAT 1003  
350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
1002 CTTGAGTGTAACTAGACAGCTATTTCAGGACACCGTTTGTCTCTACTATACACTT 943  
370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
942 CTCAGAGATCTATATTTCTGTGAACACCTGAACTCGCTCTCTCCAGATAAGCAAAA 883  
390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
882 GGAGCAAAACAGACTTTTGAAGAAATGATGAATATACATTCAGATCTGTAGTCAAGTGT 823  
410 IleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
822 ATTGGTGAAGAAACCAAGATGAAAGCATCAGACTTCTGTTTGTATGCTTACAGAACCA 763  
430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
762 GTACTCAACAGAGCTGACTTATGTTTATTTATGGACATTTGATGATGATGATGATGAT 703  
450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
702 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 655

RESULT 5

ABQ79518  
ID ABQ79518 standard; cDNA; 3145 BP.  
XX  
AC ABQ79518;  
XX

DT 25-NOV-2002 (first entry)

DE cDNA encoding a protein similar to human sorting nexin.

XX

KW Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;  
KW SSH 6; gene; human; nexin; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 128..2968  
FT /\*tag= a  
XX  
FN W0200262839-A2.  
XX  
PD 15-AUG-2002.  
XX  
PP 05-FEB-2002; 2002MO-EP01327.  
XX  
PR 07-FEB-2001; 2001EP-0200439.  
XX  
PA (UTMA-) UNIV MAASTRICHT.  
XX  
PI Daemen MZAP, Cleutjens CBJM, Zaman GJR;  
XX  
DR WPI; 2002-643400/69.  
XX  
DR P-PSDB; ABB81193.

Use of a polynucleotide differentially expressed in ruptured and stable atherosclerotic plaques as a marker for atherosclerosis, useful in treating, diagnosing or preventing atherosclerosis

Example 6; Page 35-37; 44pp; English.

The invention relates to the use of a polynucleotide differentially expressed in ruptured and stable atherosclerotic plaques as a marker for atherosclerosis, where the polynucleotides can be selected from the sequences shown in ABQ79517-19. The polynucleotides are useful as a marker of atherosclerosis, which may be used: (i) in the diagnosis, prevention and treatment of atherosclerosis; (ii) as serum/plasma markers to screen patients at risk for plaque instability to evaluate the effects of other treatments; (iii) in the preparation of vector molecules for the expression of the encoded protein in host cells; and (iv) in the identification of functional targets or analogues of the gene. The polynucleotides, the encoded proteins or antibodies against the proteins may be used to target other therapeutics to an unstable plaque. Modulation of the expression of the polynucleotide can increase plaque stability and therefore inhibit the progression of atherosclerotic cardiovascular disease. Modulators may be used to prepare pharmaceuticals for atherosclerotic disorders. The present sequence represents a cDNA encoding a protein similar to human sorting nexin.

SQ Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 3145  
Score: 330.00 Matches: 430  
Percent Similarity: 99.54% Conservative: 0  
Best Local Similarity: 99.54% Mismatches: 1  
Query Match: 70.97% Indels: 2  
DB: 24 Gaps: 0

US-09-744-313A-1 (1-465) x ABQ79518 (1-3145)

QY 35 SerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspAsp 54  
Db 1673 AGTACCACATCGAGGGAGCTATGTCCTAAATATGCTAGCTGAAGTGAAGATGAT 1732

QY 55 PheIleGluGluGlyIleValValMetGluAspAspSerProValGluAlaValSerThr 74  
Db 1733 TTTATTGAAGAGGTATTGTTGTAATGAGATGATCTTCAGTGGAGGCTGTGAGCACA 1792

QY 75 ProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPhePhe 94  
Db 1793 CCTAATACCTCCCGAAACCTTGCTGCATGGAAATAGCATTCATATAGTATGATCTTTT 1852

QY 95 GluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAsp 114



Db	2932	ACAAAGGAAGTACCTCTGTGACATCTTGATG	2965
RESULT 6			
AAI59030			
ID	AAI59030	standard; cDNA; 1551 BP.	
XX			
XX	AAI59030;		
AC			
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE	Human polynucleotide SEQ ID NO 1233.		
XX			
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia; ss.		
XX			
OS	Homo sapiens.		
XX			
FN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
FA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPI; 2001-442253/47.		
DR	P-PSDB; AAM39874.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Claim 1; SEQ ID NO 1233; 10078pp; English.		
XX			
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
SQ	Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;		
Alignment Scores:			
Pred. No.:	1.78e-307	Length:	1551
Score:	307.00	Matches:	307
Percent Similarity:	100.00%	Conservative:	0



US-09-744-313A-1 (1-465) x AAZ17379 (1-779)

QY 263 LeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLys 282

Db 80 CTGACCATCTCTCAGTCTCTACTTTCAGAAACCAACAGAGCTTTTCAATGATCTGTTTAA 139

QY 283 AsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnThrPheMetGlu 302

Db 140 AATAATGCAACCGTCTCTGAAATATACAGAGAGAAACCAAAATCAGAAATATTTATGGAG 199

QY 303 ValMetThrValGluGluValTyrrAspTyrrLeuMetTyrrValGlyArgValValPheGln 322

Db 200 GTGATGACTGTGAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGGTATTTTCCAG 259

QY 323 ValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu 342

Db 260 GTTCTGATGGCTTCATCATCTCTTAATGGGACCTGATCTCTTTTAAACACACCTG 319

QY 343 GluMetTyrrThrAspTyrrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg 362

Db 320 GAAATGTATGATGATCTATCTTCACTGTAACCTAGAACAGCTATTTCCAGGACCGT 379

QY 363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArg 382

Db 380 TTGGTCTCACTATAACACTTCTCAGAGATGCTATATCTGTGAAACACATCAACCTCGC 439

QY 383 SerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrrIle 402

Db 440 TCTCTCAGATGAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATACATT 499

QY 403 ProAspLeuLeuValLysCysIleGlyGluThrLysTyrrGluSerIleArgLeuLeu 422

Db 500 CCAGATCTGTAGTCAAGTGTATTGTGGAAGAACCAAGTATGAAAGCATCAGACTTCTG 559

QY 423 PheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrrValLeuLeuAspIle 442

Db 560 TTTGATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTGGACATT 619

QY 443 ValIleGlnGluLeuPhe 448

Db 620 GTGATACAGGAAGTGT 637

RESULT 8

AA02755

ID AAC02755 standard; cDNA; 358 BP.

XX AC AAC02755;

DE 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 2753.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX BP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG02749.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT

obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from

mRNAs encoding secreted proteins. An ORF has been identified within the

sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

derived from 30 different tissues. EST sequences usually correspond

mainly to the 3' untranslated region (UTR) of the mRNA because they are

often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

well suited for isolating cDNA sequences derived from the 5' ends of

mRNAs and even in those cases where longer cDNA sequences have been

obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

mRNAs with intact 5' ends and can therefore be used to obtain full length

cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

gene therapy and chromosome mapping procedures. They are used to obtain

upstream regulatory sequences and to design expression and secretion

vectors.

Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;

Alignment Scores:

Pred. No.: 1.46e-113 Length: 358

Score: 119.00 Matches: 119

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 25.59% Indels: 0

DB: 21 Gaps: 0

US-09-744-313A-1 (1-465) x AAC02755 (1-358)

QY 216 LeuProAspValAsnLeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGlu 235

Db 2 CTACCAATGTAATCTTGGGAAATATATAATCTGTTCTCGAAACCTAATGAAAGAG 61

QY 236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLys 255

Db 62 AAAGTTCAGCATTTGGAACTTTTATCATCAATTCATTAAATCTTGTGAGTCTCCAAAG 121

QY 256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLys 275

Db 122 CCTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAGAAACCAACAAAGAG 181

QY 276 LeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295

Db 182 CTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATATACAGAGAGAAACCA 241

QY 296 AsnGlnAsnTyrrPheMetGluValMetThrValGluGlyValTyrrAspTyrrLeuMetTyrr 315

Db 242 AATCAGAAATATTTATGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 301

QY 316 ValGlyArgValValPheGlnValProAspTrpIleHisLeuLeuMetGlyThr 334

Db 302 GTAGACGGGTAGTTTTCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAACC 358

RESULT 9

ABX50426

ID ABX50426 standard; cDNA; 305 BP.

XX AC ABX50426;

XX 25-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #355.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;

KW muscle deposition; fat deposition; genome mapping; gene identification;

XX gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137160-A1.

PT

XX PD 26-SEP-2002.  
XX PF 26-OCT-2001; 2001US-0983965.  
XX PR 17-DEC-1998; 98US-113678P.  
XX PR 15-DEC-1999; 99US-0465231.  
XX (BYAT// BYATT J C.  
XX (MATH// MATHIALAGAN N.  
XX (TAON// TAO N.  
XX (WARR// WARREN W C.  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-102386/09.  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
XX identification and analysis, cattle breeding or preparation of  
XX constructs for cattle gene expression and genetically improved cattle -  
XX Claim 2; SEQ ID No 355; 38pp; English.  
XX The invention relates to a purified nucleic acid molecule associated with  
XX lactation or muscle and fat deposition (designated LMFD), derived  
XX from cattle, and the LMFD nucleic acid can specifically hybridise to a  
XX second nucleic acid molecule comprising any of 5912 nucleotide  
XX sequences, appearing as ABX50072-ABX55983, or complements of them.  
XX Also included are; (1) a transformed cell having a nucleic acid  
XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
XX translated sequence that functions in the cell to cause termination of  
XX transcription and addition of polyadenylated ribonucleotides to a 3' end  
XX of the mRNA molecule; and (2) determining a level or pattern of a  
XX molecule in a bovine cell or tissue comprising: (a) incubating a marker  
XX nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
XX complement or fragment) with a complementary nucleic acid molecule  
XX obtained from the bovine cell or tissue, where hybridisation between the  
XX marker nucleic acid and the complementary nucleic acid permits the  
XX detection of the molecule; and (b) detecting the level or pattern of the  
XX complementary nucleic acid, where the detection of the complementary  
XX nucleic acid is predictive of the level or pattern of the molecule.  
XX The LMFD nucleic acid is used for determining a level or pattern  
XX of a molecule in a bovine cell or tissue. It is useful for genome  
XX mapping, gene identification and analysis, cattle breeding, preparation  
XX of constructs for use in cattle gene expression, or for genetically  
XX improving cattle. The present sequence is one of the 5912 bovine  
XX LMFD EST (expressed sequence tag) nucleic acids.  
XX Note: The present sequence was not shown in the specification but  
XX was obtained in electronic format from the USPTO web site:  
XX seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX SQ Sequence 305 BP; 109 A; 60 C; 51 G; 85 T; 0 other;  
Alignment Scores:  
Pred No.: 2.82e-73 Length: 305  
Score: 80.00 Matches: 80  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 17.20% Indels: 0  
DB: 25 Gaps: 0  
US-09-744-313a-1 (1-465) x ABX50426 (1-305)  
Qy 180 GlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAla 199  
Db 1 CAGGAATATCTGCAGAAACTCTTCGACATCCAGATTGAGTAACAGTCAACTCTCGGCA 60  
Qy 200 AspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysLeuProAspVal 219  
Db 61 GATTTCCTCTCCCAATGTGTGGGAAACACAGTTCTTCATAGATACATACAGACGTA 120  
Qy 220 AsnLeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGlyGlnHis 239

Db 121 AATCTGGGAAATTTATAAATCTGTCTCTGGAATACTAATGAAGAGAAAGGTCAACAT 180  
Qy 240 LeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSer 259  
Db 181 TTGGAACTTTCATCATGATTTTCAATTAATCTTGTGGAATCTCCAAAGCTTAACCGAGT 240  
RESULT 10  
ID ABX50611 standard; cDNA; 402 BP.  
XX AC ABX50611;  
XX DT 25-FEB-2003 (first entry)  
XX DE Bovine EST associated with lactation/muscle/fat deposition #540.  
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
XX KW muscle deposition; fat deposition; genome mapping; gene identification;  
XX KW gene analysis; cattle breeding.  
XX OS Bos Taurus.  
XX PN US2002137160-A1.  
XX PD 26-SEP-2002.  
XX PF 26-OCT-2001; 2001US-0983965.  
XX PR 17-DEC-1998; 98US-113678P.  
XX PR 15-DEC-1999; 99US-0465231.  
XX (BYAT// BYATT J C.  
XX (MATH// MATHIALAGAN N.  
XX (TAON// TAO N.  
XX (WARR// WARREN W C.  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-102386/09.  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
XX identification and analysis, cattle breeding or preparation of  
XX constructs for cattle gene expression and genetically improved cattle -  
XX Claim 2; SEQ ID No 540; 38pp; English.  
XX The invention relates to a purified nucleic acid molecule associated with  
XX lactation or muscle and fat deposition (designated LMFD), derived  
XX from cattle, and the LMFD nucleic acid can specifically hybridise to a  
XX second nucleic acid molecule comprising any of 5912 nucleotide  
XX sequences, appearing as ABX50072-ABX55983, or complements of them.  
XX Also included are; (1) a transformed cell having a nucleic acid  
XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
XX translated sequence that functions in the cell to cause termination of  
XX transcription and addition of polyadenylated ribonucleotides to a 3' end  
XX of the mRNA molecule; and (2) determining a level or pattern of a  
XX molecule in a bovine cell or tissue comprising: (a) incubating a marker  
XX nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
XX complement or fragment) with a complementary nucleic acid molecule  
XX obtained from the bovine cell or tissue, where hybridisation between the  
XX marker nucleic acid and the complementary nucleic acid permits the  
XX detection of the molecule; and (b) detecting the level or pattern of the  
XX complementary nucleic acid, where the detection of the complementary  
XX nucleic acid is predictive of the level or pattern of the molecule.  
XX The LMFD nucleic acid is used for determining a level or pattern  
XX of a molecule in a bovine cell or tissue. It is useful for genome  
XX mapping, gene identification and analysis, cattle breeding, preparation  
XX of constructs for use in cattle gene expression, or for genetically  
XX improving cattle. The present sequence is one of the 5912 bovine  
XX LMFD EST (expressed sequence tag) nucleic acids.  
XX Note: The present sequence was not shown in the specification but  
XX was obtained in electronic format from the USPTO web site:  
XX seqdata.uspto.gov/sequence.html?DocID=20020137160.

XX SQ Sequence 402 BP; 141 A; 71 C; 76 G; 113 T; 1 other;

Alignment Scores:

Pred. No.:	2-7e-47	Length:	402
Score:	55.00	Matches:	55
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.83%	Indels:	0
DB:	25	Gaps:	0

US-09-744-313A-1 (1-465) x ABX50611 (1-402)

QY 205 AsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIle 224

Db 3 AATGGTGGGAAACACAGATTTCTTGATAGACTACAGACGTAATCTTGGAAATT 62

QY 225 IleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPhele 244

Db 63 ATAAATCTGTTCTCTCGAAACTTAATGAAAGGAAAGGTCAACATTTGGAACCTTTCATC 122

QY 245 MetAsnPheIleAsnSerCysGluSerProLysProLysProSer 259

Db 123 ATGAATTTCAATTAATTTCTTGTAATCTCCAAAGCCTAAACCGAGT 167

RESULT 11

AAH69552

ID ABS77486 standard; cDNA; 725 BP.

XX ABS77486;

AC ABS77486;

XX 12-DEC-2002 (first entry)

XX Frog embryonic gene sequence Q9925894.

XX Frog; ss; embryonic development; developmental disorder;

XX microarray; cell differentiation.

XX Xenopus laevis.

XX US2002081610-A1.

XX 27-JUN-2002.

XX 23-JUL-2001; 2001US-0910943.

XX 21-JUL-2000; 2000US-219658P.

XX (UYRQ ) UNIV ROCKEFELLER.

XX Hammati-Brivanlou A, Altmann CR;

XX WPI; 2002-626534/67.

XX Nucleic acid array containing Xenopus embryonic nucleic acids is useful

PT to identify genes involved in embryonic development, to identify

PT different types of embryonic cells, and to diagnose developmental

PT disorders

XX Claim 1; Page 756; 823pp; English.

XX The invention relates to a nucleic acid array, where each coordinate

CC contains a single nucleic acid species having one of 770 nucleotide

CC sequences (appearing as ABS76747-ABS77516) a of a Xenopus embryonic gene

CC product, or its complement or hybridisable fragment of not less than

CC 20 contiguous nucleotides of one of those sequences. Also included are

CC detecting differential expression of embryonic genes, comprising:

CC (a) contacting a nucleic acid array comprising genes expressed in

CC embryonic but not mature cells with nucleic acids from sample and

CC control cells; and (b) detecting differential hybridisation of nucleic

CC acids from the sample cells relative to the control cells; and detecting

CC defects in development, comprising: (a) contacting nucleic acids from

CC test cells undergoing development with a nucleic acid array of gene

CC products known to play a fundamental role in the development process; and

CC (b) detecting a difference in expression of a fundamental gene in the

CC sample cells relative to a standard. The invention is useful to identify

CC genes involved in embryonic development and related processes such as

CC cell differentiation. This would be useful for diagnosing developmental

CC disorders and for identifying different types of embryonic cells.

CC The present sequence is one of the 770 Xenopus embryonic cDNA sequences.

XX SQ Sequence 725 BP; 229 A; 135 C; 140 G; 218 T; 3 other;

Alignment Scores:

Pred. No.:	1.84e-29	Length:	725
Score:	38.00	Matches:	38
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.17%	Indels:	0
DB:	24	Gaps:	0

US-09-744-313A-1 (1-465) x ABS77486 (1-725)

QY 243 PheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGlu 262

Db 173 TTTATAATGAATTTCTATAAACTCATGTGAATCACCACCAACCAACAGTAGGCTTGAA 232

QY 263 LeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280

Db 233 CTGACTATTTAGCCCGACTTCAGAAAACAATAAAAGCTTTTAAATGATCTG 286

RESULT 12

AAH69552

ID AAH69552 standard; cDNA; 311 BP.

XX AAH69552;

XX 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 826.

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.

XX WO200142467-A2.

XX 14-JUN-2001.

XX 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

XX 21-DEC-1999; 99US-0171350.

XX 14-MAR-2000; 2000US-0189315.

XX 12-MAY-2000; 2000US-0203791.

XX 09-JUN-2000; 2000US-0210600.

XX 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer

PT and for assessing and detecting compounds for treating the cancer -

XX Claim 1; Page 243; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with

CC cervical cancer with cytostatic activity. The nucleic acids and encoded

CC polypeptides are useful: to assess if a patient is afflicted with

CC cervical cancer or has a pre-malignant condition; to monitor the

CC progression of cervical cancer or a premalignant condition in a patient;

CC and to select and/or assess the efficacy of a compound or therapy for

CC inhibiting cervical cancer in a patient. The nucleic acids may also be

```
CC useful for gene therapy.
XX
SQ Sequence 311 BP; 84 A; 67 C; 63 G; 94 T; 3 other;

Alignment Scores:
Pred. No.: 2,32e-16 Length: 311
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.38% Indels: 0
DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAH69552 (1-311)
Qy 430 ValLeuAsnLysGlnLeuThrTYrValLeuAspIleValIleGlnGluLeuPhePro 449
Db 24 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGACTGTTTCCA 83

Qy 450 GluLeuAsnLysVal 454
Db 84 GAGCTCAATAAGGTA 98

RESULT 13
AAH70887
ID AAH70887 standard; cDNA; 313 BP.
XX
AC AAH70887;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 2161.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
XX
PR 21-DEC-1999; 99US-0171350.
XX
PR 14-MAR-2000; 2000US-0189315.
XX
PR 12-MAY-2000; 2000US-0203791.
XX
PR 09-JUN-2000; 2000US-0210600.
XX
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
PS WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 456; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 313 BP; 83 A; 70 C; 63 G; 97 T; 0 other;

Alignment Scores:
Pred. No.: 2,33e-16 Length: 313
Score: 24.00 Matches: 299
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.16% Indels: 0
DB: 22 Gaps: 0

us-09-744-313a-1.rng
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.38% Indels: 0
DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAH70887 (1-313)
Qy 430 ValLeuAsnLysGlnLeuThrTYrValLeuAspIleValIleGlnGluLeuPhePro 449
Db 2 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGACTGTTTCCA 61

Qy 450 GluLeuAsnLysVal 454
Db 62 GAGCTCAATAAGGTA 76

RESULT 14
AAH72445
ID AAH72445 standard; cDNA; 299 BP.
XX
AC AAH72445;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 3719.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
XX
PR 21-DEC-1999; 99US-0171350.
XX
PR 14-MAR-2000; 2000US-0189315.
XX
PR 12-MAY-2000; 2000US-0203791.
XX
PR 09-JUN-2000; 2000US-0210600.
XX
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
PS WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 696; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 299 BP; 82 A; 64 C; 57 G; 96 T; 0 other;

Alignment Scores:
Pred. No.: 2,42e-15 Length: 299
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.16% Indels: 0
DB: 22 Gaps: 0
```

US-09-744-313A-1 (1-465) x AAH72445 (1-299)

QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
Db 2 CTCACACAGCAGCTGACTTATGTTTATTGGACATTGGATACAGGAAGTGTTCACAG 61  
QY 451 LeuAsnLysVal 454  
Db 62 CTCACATAGGTA 73

RESULT 15

AAH73007  
ID AAH73007 standard; cDNA; 299 BP.

XX AC AAH73007;

DT 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 4281.

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

PF 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210800.

PR 21-JUL-2000; 2000US-0220114.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX PT New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX Claim 1; Page 941-942; 1051pp; English.  
XX PS The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX SQ Sequence 299 BP; 82 A; 64 C; 56 G; 95 T; 2 other;

Alignment Scores:

Pred. No.:	2,42e-15	Length:	299
Score:	24.00	Matches:	24
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.16%	Indels:	0
DB:	22	Gaps:	0

US-09-744-313A-1 (1-465) x AAH73007 (1-299)

QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
Db 2 CTCACACAGCAGCTGACTTATGTTTATTGGACATTGGATACAGGAAGTGTTCACAG 61



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 06:24:53 ; Search time 2929 Seconds  
(without alignments)  
3858.512 Million cell updates/sec

Title: US-09-744-313A-1  
Perfect score: 465  
Sequence: 1-MWTHFCIFRNTQKRGSEF.....ELFPELNKVKQKVTSTWSM 465

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL-frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113612\_4600/app\_query.fasta\_1.647  
-DB=EST -QFMT=fastaf -SUFFIX=st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313@CGN\_1\_1\_2810@runat\_30012004\_113612\_4600 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

- Database :
- 1: em\_estba.\*
  - 2: em\_esthum.\*
  - 3: em\_estin.\*
  - 4: em\_estmu.\*
  - 5: em\_estov.\*
  - 6: em\_estpl.\*
  - 7: em\_estro.\*
  - 8: em\_hc.\*
  - 9: gb\_est1.\*
  - 10: gb\_est2.\*
  - 11: gb\_hc.\*
  - 12: gb\_est3.\*
  - 13: gb\_est4.\*
  - 14: gb\_est5.\*
  - 15: em\_estfun.\*
  - 16: em\_estom.\*
  - 17: em\_gss\_hum.\*
  - 18: em\_gss\_inv.\*
  - 19: em\_gss\_pln.\*
  - 20: em\_gss\_vrt.\*
  - 21: em\_gss\_fun.\*
  - 22: em\_gss\_mam.\*
  - 23: em\_gss\_mus.\*
  - 24: em\_gss\_pro.\*
  - 25: em\_gss\_rtd.\*
  - 26: em\_gss\_phg.\*
  - 27: em\_gss\_vrl.\*
  - 28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	235	50.8	1096	12	BQ053127	BQ053127 AGENCOURT
2	231	49.7	921	13	BQ30190	BQ30190 AGENCOURT
3	220	47.3	664	14	CB136167	CB136167 K-EST0198
4	208	44.7	643	14	CB114674	CB114674 K-EST0158
5	206	44.3	1105	10	BG256631	BG256631 602370809
6	202	43.4	608	12	BM829345	BM829345 K-EST0102
7	199	42.8	604	14	CB130920	CB130920 K-EST0180
8	199	42.8	606	12	BM699448	BM699448 UI-B-DX1-
c	195	41.9	917	13	BX372281	BX372281
9	194	41.7	625	10	BG356529	BG356529 602564936
11	179	38.5	1076	10	BE874550	BE874550 601489664
12	176	37.8	817	10	BF92296	BF92296 602249179
13	170	36.6	513	14	CB142612	CB142612 K-EST0196
14	168	36.1	506	12	BM835779	BM835779 K-EST0111
c	167	35.9	773	12	BM982264	BM982264 UI-CF-EN1
16	166	35.7	548	14	CB139915	CB139915 K-EST0193
17	164	35.3	494	10	BG547215	BG547215 602574627
18	164	35.3	726	10	BG432502	BG432502 602495646
19	162	34.8	638	10	AW957704	AW957704 EST369774
c	159	34.2	903	9	AL578668	AL578668 AL578668
21	157	33.8	473	14	CB135556	CB135556 K-EST0187
22	157	33.8	479	10	AW96088	AW96088 QV3-BN004
23	157	33.8	570	10	BF104913	BF104913 601822645
24	157	33.8	818	10	BQ191447	BQ191447 RST10541
25	156	33.5	952	13	BQ717610	BQ717610 AGENCOURT
26	152	32.7	497	14	CB137280	CB137280 K-EST0189
27	152	32.7	558	14	CB160118	CB160118 K-EST0219
28	151	32.5	457	10	BF978483	BF978483 602148893
29	150	32.3	717	13	BQ959874	BQ959874 AGENCOURT
30	149	32.0	559	12	BM693800	BM693800 UI-B-DWI-
c	148	31.8	1043	9	AL574298	AL574298 AL574298
32	147	31.6	629	13	BQ579763	BQ579763 im90e05.x
c	143	30.8	431	9	AA417301	AA417301 zu07h02.r
34	140	30.1	428	13	BQ580018	BQ580018 im90e05.y
35	140	30.1	430	10	BE940434	BE940434 RC3-UT006
c	135	29.0	594	13	BQ738240	BQ738240 UI-B-DWI-
37	135	29.0	640	9	AW327722	AW327722 dr01d07.x
38	135	29.0	936	10	BQ026479	BQ026479 602291871
39	134	28.8	440	9	AW796463	AW796463 CMI-DM003
c	134	28.8	598	13	BQ683958	BQ683958 UI-CF-ENO
41	132	28.4	421	9	AA431154	AA431154 zw71f09.r
42	132	28.4	439	13	BQ083624	BQ083624 K-EST0146
c	132	28.4	647	13	BQ679457	BQ679457 UI-CF-DU1
c	130	28.0	986	13	BUI68735	BUI68735 AGENCOURT
c	129	27.7	564	12	BM670092	BM670092 UI-B-DX1-

ALIGNMENTS

RESULT 1  
BQ053127  
LOCUS  
DEFINITION  
AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711  
5', mRNA sequence.  
ACCESSION  
BQ053127  
KEYWORDS  
BQ053127.1 GI:19812467  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1096)



Db	5	TTGGGAAATATATAATCTGTCCTCGAATACTAATGAAGAAAGGTCAGATTG	64
Qy	241	GluProPheIleMetAsnPhelIleAsnSerCysGluSerProLysProLysProSerArg	260
Db	65	GAACCTTTATCATGAATTCATTAATCTCTGAGTCTCCAAAGCCTAAACCAAGTAGA	124
Qy	261	ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu	280
Db	125	CCAGAACTGACCATCTCTCAGCCCTACTTTCAGAAACCAACAAGAGCTTTTCAATGATCTG	184
Qy	281	PhelYsAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyPhe	300
Db	185	TTTAAATATATGAAACCGTGCTGAAATATACAGAGAAAGCAAAATCAGAAATATTTT	244
Qy	301	MetGluValMetThrValGluGlyValTyRAspTyRLeuMetTyRValGlyArgValVal	320
Db	245	ATGGAGGTGATGACTGTAGAAGAGCTATGATTAACCTGATGATGTAGGACGGGTAGTT	304
Qy	321	PheGlnValProAspTyrLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsn	340
Db	305	TTCCAGGTTCCCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAATCCTCTTTAAAAAC	364
Qy	341	ThrLeuGluMetTyRThrAspTyRLeuGlnCysLysLeuGluGlnLeuPheGlnGlu	360
Db	365	ACCTCGGAAATGATACCTGATTAATCTTCTAGTGTAACTAGAACAGCTATTTCAGGAG	424
Qy	361	HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu	380
Db	425	CACCGTTTGTCTCACTATAACACTTCTCAGAGATGCTATATTTCTGTGAAACACATGAA	484
Qy	381	ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn	400
Db	485	CCTCGCTCTCTCAAGATAAGCAAAAGAGGAGCAAAACAGACTTTTGAAGAAATGATGAAT	544
Qy	401	TyrIleProAspLeuValLysCysIleGlyGluThrLysTyRLeuSerIleArg	420
Db	545	TACATTCAGATCTGTAGTCAAGTGTATGTGTGAAGAAACCAAGTATGAAGCATCAGA	604
Qy	421	LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyRValLeuLeu	440
Db	605	CTTCTGTTGTAGGCTTACAGAACAGTCTCAACAGCAGCTGACTGATTTGTTTATTG	664
Qy	441	AspIleValIleGlnGluLeuPheProGluLeu	451
Db	665	GACATTGTGATACAGGAAGCTGTTCCAGAGCTC	697
CB136167 664 bp mRNA linear EST 29-JAN-2003			
K-EST0188577 L14ChoICKO Homo sapiens cDNA clone L14ChoICKO-20-H05			
5', mRNA sequence.			
CB136167			
CB136167.1 GI:28103366			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 664)			
Oh,K.N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Kim,Y.S., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.			
21C Frontier Korean EST Project 2001			
Unpublished			
Contact: Kim YS			
Genome Research Center			
Korea Research Institute of Bioscience & Biotechnology			
52 Eeun-dong Yuseong-gu, Daejeon 305-333, South Korea			
Tel: +82-42-860-4470			
Fax: +82-42-860-4409			
Email: yongseung@mail.kribb.re.kr			
Plate: 20 row: H column: 05			
RESULT 3			
CB136167			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
D AUTHORS			
TITLE			
JOURNAL			
COMMENT			
High quality sequence stop: 664.			
Location/Qualifiers			
1..664			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="L14ChoICKO-20-H05"			
/sex="M"			
/cell_line="Choi-CK"			
/lab_host="Top10P"			
/clone_lib="L14ChoICKO"			
/note="Organ: Liver; Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."			
BASE COUNT 227 a 107 c 146 g 184 t			
ORIGIN			
Alignment Scores:			
Pred. No.: 2,696-211 Length: 664			
Score: 220.00 Matches: 220			
Percent Similarity: 100.00% Conservative: 0			
Best Local Similarity: 100.00% Mismatches: 0			
Query Match: 47.31% Indels: 0			
DB: 14 Gaps: 0			
US-09-744-313A-1 (1-465) x CB136167 (1-664)			
Qy	14	GlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLysGlyValPhe	33
Db	3	CAGAAAGGGGAGAAATCATTTGGATTCAGCAAGATAGGTAGCAAAATTAAGAGGTATTC	62
Qy	34	LysSerThrThrMetGluGlyAlaMetLeuProAsnTyRgLyValalaGluGlyLysAsp	53
Db	63	AAAGTACACATCGAGGAGCTATGTTGCCTAATTAATGTTAGTCAAGGTGAAGAT	122
Qy	54	AspPheIleGluGlyIleValMetGluAspAspSerProValGluAlaValSer	73
Db	123	GAATTTATTGAAGAGGTATTTGTTAATGAAGATGATTTCTCCAGTGAGGCTGAGAC	182
Qy	74	ThrProAsnThrProArgAsnLeuAlaTyrLysIleSerIleProTyRValAspPhe	93
Db	183	ACACCTAATACTCCCGAAACCTTGCTGCTGATGAAATAGCAATTCATATGTAGACTTT	242
Qy	94	PheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCysIle	113
Db	243	TTTGAGGATCCCTCCCTCTGAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG	302
Qy	114	AspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyR	133
Db	303	GATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTCAACATTGGTCTGTCTAT	362
Qy	134	ArgArgTyRLeuGlnPheTyRValLeuGluSerLysIleThrLysPheHisGlyAlaPhe	153
Db	363	AGAAGATATCTTGAATTTCTATGATTTGAATCAAACTAACAGATTTCAATGTGATTT	422
Qy	154	ProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyRgLyLysLeuLys	173
Db	423	CCTGATCCCGAGCTTCTCTTAAGAGAGATCTTGGCCCCCAAAATTTATGATTTCTTAAG	482
Qy	174	SerLysArgGluGlnPheGlnGluTyRLeuGlnLysLeuLeuGlnHisProGluLeuSer	193
Db	483	TCAAAGAGGAGAGGATTCAGAGATATCTACAGAACTTCTGAGCATCCAGAACTGAGT	542
Qy	194	AsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGlnThrGlnPheLeuAsp	213
Db	543	AATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGGTGGGAGAAACACAAATTTCTTGAT	602
Qy	214	LysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMet	233
Db	603	AAGATATACAGATGTAAATCTTGGGAAATTTATAAAATCTGTCTCTGGAACTAATG	662

RESULT 4  
CB114674  
LOCUS  
DEFINITION K-EST0158382 L10choCK0 Homo sapiens cDNA clone L10choCK0-1-D02 5',  
mRNA sequence.  
ACCESSION CB114674  
VERSION CB114674.1 GI:27940481  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 643)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 1 row: D column: 02  
High quality sequence stop: 643.

FEATURES  
Location/Qualifiers  
1..643  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L10choCK0-1-D02"  
/sex="M"  
/cell\_line="Cho-CK"  
/lab\_host="Top10F"  
/notes="Organ: Liver; Vector: pVT3-Pac; Site 1: EcoRI;  
Site 2: NotI; The library was contributed by the Soares  
laboratory and it was constructed as described by Bonaldo,  
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research  
6(9): 791-806. RNA was prepared from harvested cell  
culture."

BASE COUNT 214 a 123 c 126 g 180 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3,33e-199 Length: 643  
Score: 208.00 Matches: 208  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 44.73% Indels: 0  
DB: 14 Gaps: 0

US-09-744-313A-1 (1-465) x CB114674 (1-643)

Qy 258 ProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPhe 277  
Db 1 CCAAGTCGACCAAGTACCATCTCTCAGCCCTTCTCAGAAACCAAGAGGCTTTTC 60  
Qy 278 AsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGln 297  
Db 61 AATGATCTGTTTAAATATATGAAACCGTCTGAAATACAGAGAGAAAGCAATACAG 120  
Qy 298 AsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGly 317  
Db 121 AATTATTTTATGAGGGTGATGATCTGTAGAGGAGTCTATGATTACCTGATGTATGAGA 180  
Qy 318 ArgValValPheGlnValProAspTyrPheHisLeuLeuMetGlyThrArgLysLeu 337  
Db 181 CGGGTAGTTTTCCAGGTTCTGACTGCTTCATCATCTCTTAATGGGAAGTCGAATCCTC 240

Qy 338 PheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeu 357  
Db 241 TTTTAAAAACACCTCGAAATGTATCTGATTACTATCTTCACTGTAATCAAGCAGCTA 300  
Qy 358 PheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGlu 377  
Db 301 TTTTCAGGAGCAGCGCTTGGTCTCACTCAATCACTTCTCAGAGATGCTATATCTGTGAA 360  
Qy 378 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGlu 397  
Db 361 AACACTGAACCTCGCTCTCTCCAAAGATAAGCAAAAGAGGAGCAACAGACTTTTGAAGAA 420  
Qy 398 MetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGlu 417  
Db 421 ATGATGAATTACATTCAGATCTGTGTAGTCAGTGTATTTGGTGAAGAAACCAAGTATGAA 480  
Qy 418 SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyr 437  
Db 481 AGCATCAGACTTCTGTTTGATGGCTTACAGCAACAGTACTCAACAGCAGCTGACTTAT 540  
Qy 438 ValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGlu 457  
Db 541 GTTTTATGGACATTTGTGATACAGGAACCTGTTTCCAGAGCTCAATAAGGTACAAAAGGAA 600  
Qy 458 ValThrSerValThrSerTyrMet 465  
Db 601 GTTACCTCTGTGACATCTTGGATG 624

RESULT 5  
BG256631 1105 bp mRNA linear EST 13-PBB-2001  
LOCUS 602370809F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4478958 5',  
mRNA sequence.  
ACCESSION BG256631  
VERSION BG256631.1 GI:12766447  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1105)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10310 row: a column: 03  
High quality sequence stop: 645.

FEATURES  
Location/Qualifiers  
1..1105  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4478958"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH MGC Library."

BASE COUNT 380 a 235 c 208 g 282 t  
ORIGIN

Alignment Scores:

```

Pred. No.: 6,18e-197 Length: 1105
Score: 206.00 Matches: 222
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 1
Query Match: 44.30% Indels: 1
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x BG256631 (1-1105)

QY 217 ProAspValAsnLeuGluValPheLeuMetAsnPheLeuSerValProGluValMetLysGluLys 236
Db 1 CCAGATGTAATCTCGGAAATATATAATCTGTTCTCGAAATCTATGAAAGAGAAA 60

QY 237 GlyGlnHisLeuGluProPheLeuMetAsnPheLeuSerCysGluSerProLysPro 256
Db 61 GGTGACATTTGAACTTTTATCATGATTTTCAATTAATCTTGAGCTCTCAAGCT 120

QY 257 LysProSerArgProGluLeuThrLysLeuSerProThrSerGluAsnLysLysLeu 276
Db 121 AAACCAAGTAGACCAAGACTGACCACTTCAGCCCTACTTCAGAAACCAACAGAGCTT 180

QY 277 PheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsn 296
Db 181 TTCAATGATCTGTTTAAATAATATGCAACCGCTGCTGAAATATACAGAGAGCAAAAT 240

QY 297 GlnAsnTyrPheMetGluValMetThrValGluGluValTyrAspTyrLeuMetTyrVal 316
Db 241 CAGATTTATTTATGAGGTGATGATCTAGAGAGTCTATGATTAATCTGATGATGTA 300

QY 317 GlyArgValValPheGlnValProAspTyrLeuHisLeuMetGlyThrArgile 336
Db 301 GGAAGGTAGTTTCCAGGTTCTGACTGCTTCATCACTCTTATGGAAGTGAATC 360

QY 337 LeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGln 356
Db 361 CTCTTTAAACACCTCGAAATGATATCTGATTAATCTTCCAGTGAATCTAGAACAG 420

QY 357 LeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaLeuPheCys 376
Db 421 CTATTTCCAGGAGCACCGTTTGGTCTCACTCATCAACACTTCTCAGAGATCTATCTGT 480

QY 377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396
Db 481 GAAACACTGAACTGCTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTGA 540

QY 397 GluMetMetAsnTyrLeuProAspLeuValLysCysLeuGluThrLysTyr 416
Db 541 GAAATGATGAATACATTCAGATCTGTTAGTCAAGTATGTTGTTGAGAAACCAAGTAT 600

QY 417 GluSerLeuArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436
Db 601 GAAAGCATCAGACTTCTGTT-GATGGCTTACAGCAACCACTACTCAACAGCAGCTGACT 659

QY 437 TyrValLeu 439
Db 660 TATGTTTATA 668

RESULT 6
BM829345
LOCUS X-EST0102319 S9SNU601 Homo sapiens cDNA clone S9SNU601-45-A08 5',
DEFINITION mRNA sequence.
ACCESSION BM829345
VERSION 1 GI:19185754
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 608)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

```

TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongseung@mail.kribb.re.kr  
Plate: 45 row: A column: 08  
High quality sequence stop: 608.  
Location/Qualifiers  
1. 608

## FEATURES

source  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S9SNU601-45-A08"  
/sex="M"  
/tissue\_type="Ascites"  
/cell\_type="Epithelial"  
/cell\_line="SNU-601"  
/lab\_host="Top10F"  
/clone\_lib="S9SNU601"  
/notes="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;  
Site 2: XhoI; The poly (A) RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then dephosphorylated  
with tobacco acid pyrophosphatase (TAP). The dephosphorylated  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DraIII-digested pME18S-FL3 vector. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."  
BASE COUNT 203 a 100 c 130 g 175 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3,56e-193 Length: 608  
Score: 202.00 Matches: 202  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.44% Indels: 0  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x BM829345 (1-608)

QY 27 SerLysLeuLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyr 46  
Db 2 AGCAAAATTAAGAGATATTCAAAAGTACCACATGAGGAGCTATGTTGCCCTAATAT 61  
QY 47 GlyValAlaGluGlyGluAspAspPheLeuGluGlyLeuValMetGluAspAsp 66  
Db 62 GGTGTAGCTGAAGGTGAAGATGATTTATTGAAGAGATATTGTTGAATGAGAGATGAT 121  
QY 67 SerProValGluAlaValSerThrProAsnThrProArgAsnLeuAlaLysLysLeu 86  
Db 122 TCTCCAGTGGAGGTGTGAGCACACTTACTCCCGAAACCTTGTGCTGATGAAAT 181  
QY 87 SerLeuProTyrValAspPhePheGluAspProSerSerGluArgLysGluLysGlu 106  
Db 182 AGCATTCATATAGACTTTTTCAGGATCCCTCTGAAAGAGAGAGAGAGAGAA 241  
QY 107 ArgLeuProValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGlu 126  
Db 242 AGAATTCCTGTTGTTTATGATGTTTCAAGAGAAATGATAGAGAGAGAGAGAGAG 301  
QY 127 ProGluHisTyrSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeu 146

Db 302 CTGAACATGGTCTCTATAGACATATCTTGAATCTTATGACTTGAATCAAACTA 361  
Qy 147 ThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleleGlyPro 166  
Db 362 ACAGAAATTCATGGTGCATTTCTGATGCCAGCTTCTTCTTAAGAGGATCATTTGCCCC 421  
Qy 167 LysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeu 186  
Db 422 AAAAAATATGAATTTCTTAAGTCAAGAGGAGAGAGTTCCTCAAGAAATATATACAGAACTT 481  
Qy 187 LeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGly 206  
Db 482 CTGAGCATCCAGAACTAGATATAGTCACTTCTGCGAGACTTTCTTCCCTTAATGT 541  
Qy 207 GlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleleLys 226  
Db 542 GGGGAACACAAATTTCTTGATAGATACACAGATGTAAATCTTGGGAATTTATAAA 601  
Qy 227 SerVal 228  
Db 602 TCTGTT 607

RESULT 7  
CB130920 604 bp mRNA linear EST 29-JAN-2003  
LOCUS K-EST0180911 L12JSHC0 Homo sapiens cDNA clone L12JSHC0-5-A04 5',  
DEFINITION mRNA sequence.  
ACCESSION CB130920.1 GI:28095369  
VERSION CB130920.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 604)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 5 row: A column: 04  
High quality sequence stop: 604.  
Location/Qualifiers

FEATURES  
source 1..604  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L12JSHC0-5-A04"  
/sex="M"  
/cell\_line="J-SHC"  
/lab\_host="Top10P"  
/clone\_lib="L12JSHC0"  
/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;  
Site 2: NotI; The poly (A) + RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including  
EcoRI site by treatment of T4 RNA ligase and the first  
strand cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10P by electroporation method.

The cDNA libraries constructed by this method are  
full-length enriched cDNA library."  
BASE COUNT 205 a 100 c 137 g 162 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3-76e-190 Length: 604  
Score: 199.00 Matches: 199  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 42.80% Indels: 0  
DB: 14 Gaps: 0

US-09-744-313A-1 (1-465) x CB130920 (1-604)  
Qy 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyLysSerArgIleleGlySerLysIleLys 30  
Db 7 AGGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAATATAA 66  
Qy 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50  
Db 67 GGAGTATTTCAAAAGTACCACAAATGGAGGAGCTATGTTGCTTAATATGTTAGTCTGAA 126  
Qy 51 GlyGluAspAspPheIleGluGlyValValMetGluAspAspSerProValGlu 70  
Db 127 GGTGAAGATGATTTTATTTAGAGAGGTATTTGTAATGAGAGATGATTTCTCCAGTGGAG 186  
Qy 71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyr 90  
Db 187 GCTGTGAGCACACCTAATACTCCCGAAACCTTGCTGATCGTGAATAATAGCATTCATAT 246  
Qy 91 ValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProVal 110  
Db 247 CTAGACTTTTTTGGAGTATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGT 306  
Qy 111 PheCysIleAspValGluArgAsnAspArgArgAlaValGlyHisGluProGluHisTrp 130  
Db 307 TTTTGTATGATGTTGAAAGAAATAGTAGAGACAGTGTGACACAGAGCTGACATGG 366  
Qy 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
Db 367 TCTGTCTATAGAGATATCTTGAATTTCTGATCTTCAATCAAACTAAACAGATTTTCA 426  
Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleleGlyProLysAsnTyrGlu 170  
Db 427 GGTGCTATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAATATGAA 486  
Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisPro 190  
Db 487 TTTTAAAGTCAAAAGAGGAGAGAGTTCCTCAAGAAATATCTACAGAACTTCTGCGAGCTCA 546  
Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 547 GAACGTAGTAATAGTCAACTTCTGCGAGACTTTCTTCTTCCCTATGTTGGGGAACA 603

RESULT 8  
LOCUS BM699448 606 bp mRNA linear EST 28-FEB-2002  
DEFINITION UI-E-DX1-agg-g-17-0-UI.r1 UI-E-DX1 Homo sapiens cDNA clone  
UI-E-DX1-agg-g-17-0-UI.5', mRNA sequence.  
ACCESSION BM699448  
VERSION BM699448.1  
KEYWORDS EST. GI:19012706  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 606)  
AUTHORS Bonaldo,M.P., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477

889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEHRP, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@iowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .606  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DX1-agv-g-17-0-UI"  
/tissue\_type="fetal eyes"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DX1"  
/notes="Organ: eye; Vector: p773-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-DX1 is a normalized cDNA library containing the  
following tissue(s): fetal eyes. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into p773-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
AGAAATCAAGA. This library was created for the program, Gene  
Discovery in the Visual System, supported by National Eye  
Institute (NEI)."  
204 a 114 c 118 g 170 t  
BASE COUNT  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,77e-190 Length: 606  
Score: 199.00 Matches: 199  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 42.80% Indels: 0  
DB: 12 Gaps: 0  
US-09-744-313a-1 (1-465) x BM699448 (1-606)  
QY 267 SerProThrSerGluAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsn 286  
DB 2 AGCCTACTTCAGAAACACACAGAGCTTTTCATGATCTGTTAAATATATGCAAC 61  
QY 287 ArgAlaGluAsnThrGluArgLysGluAsnGlnAsnTyrPheMetGluValMetThrVal 306  
DB 62 CGTGCTGAATATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGATGATGCTGTA 121  
QY 307 GluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyr 326  
DB 122 GAAGGAGTCTATGATTAATCTGATGATGATGAGAGCGGTAGTTTCCAGGTTCCGACTGG 181  
QY 327 LeuHisLysLeuLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThr 346  
DB 182 CTTTCATCATCTCTTAATGGGAACTCGAATCTCTTTTAAACACACCCCTGGAAATGATPACT 241  
QY 347 AspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeu 366  
DB 242 GATTACTATCTTCAGTAACTAGACAGCTATTTTCAGGAGCACCGTTGGTCTCACTC 301

QY 367 IleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAsp 386  
DB 302 ATACACTTCTCAGAGATGCTATATCTTGTAAGAAACACTGAACTGCTCTCTCCAGAT 361  
QY 387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
DB 362 AAGCAAAAGAGCAAAACAGACTTTTGAAGAATGATGATATACATTCAGATCTGTTA 421  
QY 407 ValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeu 426  
DB 422 GTCAAGTGATTTGTTGAAGAAACCAAGATATCAAGAGCATCAGACTTCTGTTGATGGCTTA 481  
QY 427 GlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGlu 446  
DB 482 CAGCAACAGTACTCAACAGAGCTGACTATGTTTATTTGACATTTGATACAGAA 541  
QY 447 LeuPheProGluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
DB 542 CTGTTTCAGAGCTCAATATAGTACAAAGAGATTAACCTCTCTGTGACATCTTGGATG 598  
RESULT 9  
BX372281/c  
LOCUS BX372281 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL007YE10 3-PRIME, mRNA sequence.  
ACCESSION BX372281  
VERSION BX372281  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 917)  
Li W.B., Gruber C., Jesses, J. and Polayes D.  
Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 3600.r for  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAI044ZB11\_CS04210\_1&cluster=3600.r  
Contact : Peng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BAI044ZB11\_CS04210\_1.  
Location/Qualifiers  
1. .917  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL007YE10"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_lines="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 257 a 168 c 177 g 310 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6,38e-186 Length: 917  
Score: 195.00 Matches: 253  
Percent Similarity: 98.83% Conservative: 0  
Best Local Similarity: 98.83% Mismatches: 1  
Query Match: 41.94% Indels: 3  
DB: 13 Gaps: 0



US-09-744-313A-1 (1-465) x BX372281 (1-917)

QY 181 GluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
Db 777 GAATATCTACAGAACTCTCGAGCATTCAGAACTGAGTAATAGTCACTCTGGCAGAC 718  
QY 201 PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220  
Db 717 TTTCTTCCCTAATGGTGGGAAACAAATTTCTTGATAGATACCTACAGATGTAAT 658  
QY 221 LeuGlyLysIleLys-SerValProGlyLysLeuMetLysGlyGlnHisLe 240  
Db 657 CTTGGGAAATATATANA-ATCTGTCTCTGGAAACTAATGAAGAGAAAGCTCAGCATTT 599  
QY 240 uGluProPheIleMetAsnPhelAsnSerCysGluSerProLysProSerAr 260  
Db 598 GGAACCTTTTATCATGAATTTTCAATTAATCTTGAGTCTCCAAAGCTTAAACCAAGTAG 539  
QY 260 GProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLe 280  
Db 538 ACAGAGACTGACATCTCTCAGCCCTACTTCAGAAACACACAGAGAGCTTTTCAATGATCT 479  
QY 280 uPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPh 300  
Db 478 GTTTAAATAATATGCAACCGTCTGAAATATACAGAGAGAAAGCAAAATCAGAAATATT 419  
QY 300 eMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVa 320  
Db 418 TATGGAGGTGATGACTGTAGAAGAGTCTATGATTAACCTGATGTATGAGACGGGTAGT 359  
QY 320 lPheGlnValProAspTyrLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAs 340  
Db 358 TTTCCAGGTTCTGAGTCTCATCATCTTATATGGAACTCGATCTCTTTAAAAA 299  
QY 340 nThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGl 360  
Db 298 CACCTCGGAAATGTAATGATTAATCTTCTGAGTAACTAGAACAGCTATTTTCAGGA 239  
QY 360 uHisArgLeuValSerLeuThrLeuLeuHisLeuMetGlyThrArgIleLeuPheLysAs 380  
Db 238 GCACCGTTTGGTCTCACTCATACACTTCTCAGAGATGCTATATTTCTGTGAAACACTGA 179  
QY 380 uProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAs 400  
Db 178 ACCTCGCTCTCTCCAGATGAAGCAAAAGAGCAAAACAGACTTTTGAAGAAATGATGA 119  
QY 400 nTyrIleProAspLeuValLysCysIleGlyGluGluThr-LysTyrGluSerIleA 420  
Db 118 TTACATCCAGATCTGTAGTCAAGTGTATTGTTGTAAGAAACNCAAGATGAAGCAICA 59  
QY 420 rgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGln 434  
Db 58 GACTTCTGTTTGTGGCTTACAGCAACCACTACTCAACAAGCAG 15

RESULT 10  
BG536529 625 bp mRNA linear EST 03-APR-2001  
LOCUS 602564936F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4689464 5',  
DEFINITION mRNA sequence.

ACCESSION BG536529  
VERSION BG536529.1 GI:13528075  
KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 625)

NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapsb-re@mail.nih.gov](mailto:cgapsb-re@mail.nih.gov)

Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: L16M505 row: h column: 09  
High quality sequence stop: 625.

## FEATURES

source

1..625  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4689464"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH MGC 77"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:  
SfiI (ggccgctcgcc); Site 2: SfiI (ggccattagcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGGCGGCGCATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.9  
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH MGC Library."

BASE COUNT 211 a 123 c 119 g 172 t

## ORIGIN

Alignment Scores:  
Pred. No.: 4,33e-185 Length: 625  
Score: 194.00 Matches: 194  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 41.72% Indels: 0  
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x BG536529 (1-625)

QY 248 lLeAsnSerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSer 267  
Db 2 ATTAATTTCTGTGAGTCTCCAAAGCTTAACCAAGTAGACCAAGTACCATTCTCAGC 61  
QY 268 ProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArg 287  
Db 62 CCTACTTCAGAAACACACAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGT 121  
QY 288 AlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGlu 307  
Db 122 GCTGAAATACAGAGAGAGAAAGCAAAATCAGAAATTAATTTATGAGAGTGTACTGTAGAA 181  
QY 308 GlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeu 327  
Db 182 GGAGTCTATGATTAACCTGATGTATGAGAGCGGTAGTTTTCAGGTTCTCTGACTGGCTT 241  
QY 328 HisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAsp 347  
Db 242 CATCATCTCTTAATGGGAACCTCGAATCTCTTTAAACACACCTTGGAAATGTATATCAT 301  
QY 348 TyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlnHisArgLeuValSerLeuIle 367  
Db 302 TACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCCTACTATA 361  
QY 368 ThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLys 387  
Db 362 ACACCTTCTCAGAGATGCTATATTTCTGTGAAACACACTGAACTCGCTCTCTCCAGATGAG 421  
QY 398 GlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuVal 407  
Db 422 CAAAAGAGGCAAAACAGACTTTTGAAGAAATGATGAATTACATTCCAGATCTGTTAGTC 481  
QY 408 LysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGln 427

DB	FEATURES	source
482 Ddb	AAAGTGTATCGGTGAGAAACCAAGATGAAGCATCAGACTTCTGTTTGTATGGCTTACAG 541	
428 Qy	GlnProValLeuAnLysGlnLeuThrTyrrValLeuLeuasp 441	
542 Ddb	CAACAGTACTCAACAAGCAGCTGACTTATGTTTATTGAC 583	
RESULT 11		
BB874550		
LOCUS	BB874550	
DEFINITION	601489664F1 NIH_MGC 69 Homo sapiens cDNA clone IMAGE:3892095 5', mRNA linear EST 20-OCT-2000	
ACCESSION	BB874550	
VERSION	BB874550.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 1076)	
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCTD/DTF/Gazdar DNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) Cloning Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM9677 row: p column: 16 High quality sequence stop: 705. Location/Qualifiers 1. .1076 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone IMAGE:3892095" /tissue type="large cell carcinoma, undifferentiated" /lab host="DH10B (phage-resistant)" /clone lib="NIH_MGC 69" /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."	
BASE COUNT	388 a 164 c 233 g 291 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.06e-169	Length: 1076
Score:	179.00	Matches: 179
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	38.49%	Indels: 0
DB:	10	Gaps: 0
US-09-744-313A-1 (1-465) x BB874550 (1-1076)		
Qy	79 ArgAsnLeuAlaAlaTrpLysIleSerIleProTyrrValAspPhePheGluAspProSer 98	
Ddb	2 CGAAACCTTGCTGCATGGAATTAAGCATTCATATGAGCTTTTGGAGTCCCTCC 61	
Qy	99 SerGluArgLysGluLysGluArgIleProValPheCysIleAspValGluArgAsn 118	
Ddb	62 TCTGAAGCAGGAGGAGAAAAAGAAATTCCTGTGTTTGTATGTGTAAGAAAT 121	
Qy	119 AspArgArgAlaValGluLysGluProGluHisTrpSerValTyrrArgGlyLeuGlu 138	
Ddb	122 GATAGAAGACGAGTTGGACAGAGCCTGAAATTGGTCTGTCTATAGAATATCTTGA 181	
Qy	139 PheTyrrValLeuGluSerLysLeuThrGluPheHisGluValAlaPheProAspAlaGlnLeu 158	

Alignment Scores: 8.38e-167 Length: 817  
Pred. No.: 176.00 Matches: 176  
Score: 100.00% Conservatives: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.85% Gaps: 0  
DB: 10

US-09-744-313A-1 (1-465) x BF692296 (1-817)

QY 253 SerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsn 272  
Db 2 TCTCCAAAGCCTAAACCAAGTAGACCAAGACTGACCAATCTCAGCCCTACTTCAGAAAAC 61  
QY 273 AsnLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGlu 292  
Db 62 AACAGAAGCTTTCAATGATCTGTTTAAATAATATGCAACCCGCTGGAATATACAGAG 121  
QY 293 ArgLysGlnAsnGlnAsnThrPheMetGluValMetThrValGluGlyValTyrAspTyr 312  
Db 122 AGAAGCAAAATCAGAAATTAATTTATGGAGGTGATGACTGTAGAGGAGTCTATGATTAC 181  
QY 313 LeuMetTyrValGlyArgValPheGlnValProAspTyrLeuHisLeuLeuMet 332  
Db 182 CTGATGTATGAGACGGGTAGTTTCCAGGTTCTCTGACTGCTTCATCATCTCTTAATG 241  
QY 333 GlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCys 352  
Db 242 GGAATCGAATCTCTTTAAACACCCCTGGAATATGATATGATTACTATCTTCAGTGT 301  
QY 353 LysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAsp 372  
Db 302 AAATAGAACAGCTATTTTCCAGGACCGTTTGCTCTCACTATACACTTCTCAGAT 361  
QY 373 AlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLys 392  
Db 362 GCTATATCTGTGAACACTGAACCTCGCTCTCTCCAGATAAGCAAAAGAGCAAAA 421  
QY 393 GlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGlu 412  
Db 422 CAGACTTTTGAAGAAATGATGAATATCAATTCAGATCTGTAGTCAAGTGTATGTTGTA 481  
QY 413 GluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGln 428  
Db 482 GAAACCAAGATGAAAGCATCAGACTCTGTGTTGAGCTTACAGCAA 529

RESULT 13  
CB142612 513 bp mRNA linear EST 29-JAN-2003  
LOCUS K-EST0196459 L16HLK3 Homo sapiens cDNA clone L16HLK3-25-H03 5',  
DEFINITION mRNA sequence.  
ACCESSION CB142612  
VERSION 1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM  
Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 513)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
21C Frontier  
Unpublished  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 25 row: H column: 03  
High quality sequence stop: 513.

FEATURES  
source  
1..513  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L16HLK3-25-H03"  
/cell\_line="HLK-3"  
/lab\_host="Top10F"  
/clone\_lib="L16HLK3"  
/notes="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;  
Site 2: NotI; The library was contributed by the Soares  
laboratory and it was constructed as described by Bonaldo,  
M.P., Lennon, G. and Soares, M.B. (1996), Genome Research  
6(9): 791-806. RNA was prepared from harvested cell  
culture."

BASE COUNT 165 a 89 c 112 g 147 t  
ORIGIN

Alignment Scores: 5.78e-161 Length: 513  
Pred. No.: 170.00 Matches: 170  
Score: 100.00% Conservatives: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 36.56% Gaps: 0  
DB: 14

US-09-744-313A-1 (1-465) x CB142612 (1-513)

QY 34 LysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAsp 53  
Db 3 AAAAGTACCAATGAGGAGGAGCTATGTTGCCCTAAATTATGGTGTAGCTGAAGGTGAAGAT 62  
QY 54 AspPheIleGluGluGlyIleValValMetGluAspSerProValGluAlaValSer 73  
Db 63 GATTTTATTGAAGAGGTATGTTGTAATGGAAGATGATCTCCAGTGGAGGCTGTGAGC 122  
QY 74 ThrProAsnThrProArgAsnLeuAlaIleTyrLysIleSerIleProTyrValAspPhe 93  
Db 123 ACACCTAATCTCCCGAACCTTCTGTCATGCAATAGCAATTTAGCATTCATATGTAGACTTT 182  
QY 94 PheGluAspProSerSerGluArgLysGluLysGluArgIleProValPheCysIle 113  
Db 183 TTTGAGGATCCCTCTCTGAAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242  
QY 114 AspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyrSerValTyr 133  
Db 243 GATGTTCAAGAAATATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302  
QY 134 ArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPhe 153  
Db 303 AGAAGATATCTTGATTTCTATGTTGATCAAACTAACAGATTTTCATGTTGCTAT 362  
QY 154 ProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLys 173  
Db 363 CCTGATGCCAGCTTCCTTCTTAAGAGGATCATTTGGCCCCCAAAAATTTATGATTTCTTAAG 422  
QY 174 SerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisProGluLeuSer 193  
Db 423 TCAAGAGGAGAGAGAGTTCCAGGAATATCTACAGAACTTCTGAGCACTTCAGCACTCAGT 482  
QY 194 AsnSerGlnLeuLeuAlaAspPheLeuSer 203  
Db 483 AATAGTCAACTTCTGGCAGAGACTTCTTTCC 512

RESULT 14  
BM835779  
LOCUS K-EST0110104 S11SNUI Homo sapiens cDNA clone S11SNUI-73-D02 5',  
DEFINITION mRNA sequence.  
ACCESSION BM835779  
VERSION BM835779.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 506)  
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
Oh, K.J., Cheong, J.S., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
Kim, Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongseung@mail.kribb.re.kr  
Plate: 73 row: D column: 02  
High quality sequence stop: 506.  
Location/Qualifiers  
1. .506  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S115NUL-73-D02"  
/sex="M"  
/tissue type="Stomach"  
/cell\_type="lymphoblast-like"  
/cell\_line="SNU-1"  
/lab\_host="Top10"  
/clone\_lib="S115NUL"  
/notes="Organ: Stomach; Vector: pME18-FL3; Site: 1: XhoI;  
Site: 2: XhoI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DraIII- digested pME18-FL3 vector. The  
obtained cDNA vectors were used for transfection of  
competent cells E. coli Top10<sup>+</sup> by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."  
BASE COUNT 158 a 94 c 104 g 150 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,94e-159 Length: 506  
Score: 168.00 Matches: 168  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 36.13% Indels: 0  
DB: 12 Gaps: 0  
US-09-744-313A-1 (1-465) x BM835779 (1-506)  
Qy 298 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyValAlaLysGlnThrPheGluGlu 317  
Db 2 AATATTTTATGGAGGTGATGACTGTGAGAGGAGTCTATGATACCTGATGATGATGAGGA 61  
Qy 318 ArgValValPheGlnValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeu 337  
Db 62 CGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAACCTGAATCCTC 121  
Qy 338 PheLysAsnThrLeuGluMetTyThrAspTyThrLeuGlnCysLysLeuGluGlnLeu 357  
Db 122 TTTAAACACCCCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181  
Qy 358 PheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGlu 377  
Db 182 TTTCAGGAGCACCGTTGGTCTCTACTCATACACTTCTCAGAGATGCTATATCTGTGAA 241

Qy 378 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyValAlaLysGlnThrPheGluGlu 397  
Db 242 AACACTGACCTGCTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTGAAGAA 301  
Qy 398 MetMetAsnTrpIleProAspLeuValLysCysLysGlyGluThrLysTrpGlu 417  
Db 302 ATGATGAATTACATTCAGATCTGTGTAGTCAAGTGTATTGGTGAAGAAACCAAGTATGAA 361  
Qy 418 SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTrp 437  
Db 362 AGCATCAGACTTCTGTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTAT 421  
Qy 438 ValLeuLeuAspLysValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGlu 457  
Db 422 GTTTTATTGGACATTTGTGATACAGGAACCTGTTCCAGAGCTCAATAGGTACAAAGGAA 481  
Qy 458 ValThrSerValThrSerTrpMet 465  
Db 482 GTTACCTCTGTGACATCTTGGATG 505  
RESULT 15  
BM982264/c  
LOCUS BM982264 773 bp mRNA linear EST 21-FEB-2003  
DEFINITION UI-CF-EN1-acr-j-04-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone  
UI-CF-EN1-acr-j-04-0-UI 3', mRNA sequence.  
ACCESSION BM982264  
VERSION BM982264.1 GI:19605588  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 773)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PubMed 8889548  
COMMENT Contact: McCray, PB  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.  
Location/Qualifiers  
1. .773  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-acr-j-04-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-EN1"  
/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a  
modified polylinker; Site: 1: EcoR I; Site 2: Not I;  
UI-CF-EN1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Cystic Fibrosis  
Epithelial Cells. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.

TAG LIB=UI-CF-ENI  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG SEQ=CTGCTCAGGT"

BASE COUNT 233 a 146 c 124 g 269 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 9 53e-158 Length: 773  
Score: 167.00 Matches: 183  
Percent Similarity: 99.46% Conservative: 0  
Best Local Similarity: 99.46% Mismatches: 1  
Query Match: 35.91% Indels: 1  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x BM982264 (1-773)

QY	282	LysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMet	301
DB	740	AAAAATATGCAACCGCTGTAATAATACAGAGAGAAAGCAAAATCAGAA-TATTTTATG	682
QY	302	GluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPhe	321
DB	681	GAGGTGATGACTGTAGAGAGGTCTATGATTACTGATGATGTAGGACGGGTAGTTTC	622
QY	322	GlnValProAspThrLeuHisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThr	341
DB	621	CAGGTCCCTGACTGGCTTCATCATCTCTTAATGGGAACTCGAATCTCTTTAAACACC	562
QY	342	LeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHis	361
DB	561	CTGGAATGTATCTACTGATCTTCTCAGTGTAACTAGAACAGCTATTTTCAGAGCAC	502
QY	362	ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro	381
DB	501	CGTTTGGTCTCACTATACACTTCTCAGAGATGTATATTTCTGTGAAACACCTGAACCT	442
QY	382	ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr	401
DB	441	CGCTCTCTCCAGATAGCAAAAAGGAGCAAAACACACTTTTGAGAAATGATGAATTAC	382
QY	402	IleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeu	421
DB	381	ATTCCAGATCTGTAGTCAAGTGTATTTGGTGAAGAACCAAGTATGAAGCATCAGACTT	322
QY	422	LeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAsp	441
DB	321	CTGTTTGATGCTTACAGCAACCACTACTCAACAAGCAGCTGACTTATGTTTATTGGAC	262
QY	442	IleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSerVal	461
DB	261	ATTGTGATACAGGAACGTGTTTCCAGAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTG	202
QY	462	ThrSerTrpMet	465
DB	201	ACATCTTGGATG	190

Search completed: January 31, 2004, 09:03:14  
Job time : 2945 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:37:46 ; Search time 377 Seconds  
(without alignments)  
3329.546 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLIHLCLIFRNQKRGESF.....ELPPELNKVKQKVTSTWSM 465

Scoring table

BLOSUM62  
Gapopen 10.0, Xgapext 0.5  
Ygapopen 10.0, Ygapext 0.5  
Fgapopen 6.0, Fgapext 7.0  
Delopen 6.0, Delxext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2 1/USPTO.spool/US09744313/runat 30012004 113535 4176/app\_query.fasta\_1.647  
-DB=N Geneseq 19Jun03 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 @CNG 1.0 @runat 30012004 113535 4176 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2437	100.0	1992	22	SNEXN DNA #1. Hom
2	2386	97.9	1716	22	Human polynucleoti
3	2386	97.9	2176	22	Human polynucleoti
4	2386	97.9	2176	22	Human polynucleoti
5	2382	97.7	3145	24	CDNA encoding a pr
6	2057.5	84.4	1551	22	Human polynucleoti
7	1054	43.2	779	20	Human gene express
8	981.5	40.3	725	20	Frog embryonic gen
9	682	28.0	402	25	Bovine EST associa
10	634	26.0	358	21	Human secreted pro
11	505	20.7	305	25	Bovine EST associa
12	262	10.8	1900	24	CDNA encoding huma
13	255	10.5	2133	22	Human polynucleoti
14	255	10.5	2181	22	Human polynucleoti
15	249.5	10.2	2494	24	CDNA sequence #238
16	235	9.6	2456	22	CDNA encoding nove
17	228	9.4	2869	23	DNA encoding novel
18	192	7.9	1429	24	Human secreted pro
19	192	7.9	1429	24	CDNA encoding huma
20	188	7.7	2955	22	Human cDNA sequenc
21	184.5	7.6	3585	23	DNA encoding novel
22	182.5	7.5	3435	23	Drosophila melanog
23	156	6.4	2310	24	Hypoxia-regulated
24	156	6.4	4156	21	Human ORFX ORF3067
25	155.5	6.4	2404	21	Human serum and gl
26	155.5	6.4	2483	22	Human full-length
27	155.5	6.4	2572	23	DNA encoding novel
28	155.5	6.4	2702	23	DNA encoding novel
29	155.5	6.4	2711	22	Human polynucleoti
30	155.5	6.4	2760	22	Human polynucleoti
31	155	6.4	2250	22	Novel protein kina
32	152.5	6.3	1294	23	CDNA encoding nove
33	152.5	6.3	1294	23	DNA encoding novel
34	149.5	6.1	2512	20	Human serum glucoc
35	148.5	6.1	1607	21	Human ORFX ORF2906
36	148	6.1	816	25	Aspergillus oryzae
37	147.5	6.1	1213	24	Human intracellular
38	147.5	6.1	1502	24	Human cDNA encodin
39	147.5	6.1	3106	23	Human prostate exp
40	146.5	6.0	810	22	Human cDNA 5'-end
41	146.5	6.0	810	22	Human cDNA clone r
42	142.5	5.8	311	22	Human cervical can
43	142.5	5.8	313	22	Human cervical can
44	142	5.8	995	24	Human secreted pro
45	141.5	5.8	1940	21	Human secreted pro

# ALIGNMENTS

RESULT 1  
AAC86397  
ID AAC86397 standard; DNA; 1992 BP.

AC AAC86397;

XX 21-MAR-2001 (first entry)

DT SNEXN DNA #1.

DE SNEXN; human; sorting nexin; inflammation; asthma; allergy; AIDS;  
neurological disorder; gastrointestinal; smooth muscle cell;  
cancer; Gene therapy; ds.

OS Homo sapiens.

XX WO200073334-A2.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

XX	07-DEC-2000.
PD	
XX	
XX	26-MAY-2000; 2000WO-US14831.
XX	
XX	27-MAY-1999; 99US-0136740.
PR	16-JUN-1999; 99US-0139566.
XX	(INCY-) INCYTE GENOMICS INC.
PA	
XX	Yue H, Tang YT, Azimzai Y;
XX	
DR	WPI; 2001-041141/05.
XX	
PT	Novel Sorting Nexin polypeptides and polynucleotides useful for
PT	diagnosing and treating disorders associated with their expression e.g.
PT	autoimmune disorders, smooth muscle cell disorders and cell
PT	proliferative disorders -
XX	
XX	Claim 5; Page 82-83; 84pp; English.
CC	
XX	The present invention relates to human sorting nexin (SNEKN).
CC	Compositions containing SNEKN or agonists of SNEKN are useful for
CC	treating a disease or condition associated with decreased expression of
CC	functional SNEKN and compositions containing antagonists of SNEKN are
CC	useful for treating a disease or condition associated with
CC	overexpression of functional SNEKN. These can be immune
CC	disorders for example inflammation, asthma, allergy, and AIDS,
CC	neurological disorders, gastrointestinal disorders, smooth muscle cell
CC	disorders, cancers and others. The SNEKN may also be used for somatic
CC	or germline gene therapy.
XX	
SQ	Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;
Alignment Scores:	
Pred. No.:	2,16e-232 Length: 1992
Score:	2437.00 Matches: 465
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	22 Gaps: 0
US-09-744-313A-1 (1-465) x AAC86397 (1-1992)	
Qy	1 MetTyrLeuIleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPhe 20
Dd	116 ATGTACTTGATACATTTTGGTTTAATTTCAGGAACACACAGAAAAGGGAGATCATTT 175
Qy	21 GlyIleSerArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGly 40
Dd	176 CGAAATCAGCAGAATAGGTAGCAAAAATAAAGGAGTAGTATCAAAGTAGTACCACAATGAGGGA 235
Qy	41 AlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGlyIle 60
Dd	236 GCTATGTTCCTCAATATATGGGTAGCTGAGGTGAAGTAGATGATTTTATTGAAGAAGGTATT 295
Qy	61 ValValMetGluAspAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80
Dd	296 GTTGTAAATCGAAGATGATTTCTCCAGTGAGGCTGTGAGCACACCTAATTAATCTCCCGAAC 355
Qy	81 LeuAlaAlaTrpLysIleSerIleProTyrValAspPheGluAspProSerGerGlu 100
Dd	356 CTTCGTGCATGGAATAATTAGCATTCATATGACATTTTTTTCAGGATCCCTCTCTGAA 415
Qy	101 ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120
Dd	416 AGCAAGAGAGAAAAAGAAAGAAATTCCTGTGTTTGTATTGATGTTGTAAAGAAATGATAGA 475
Qy	121 ArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyr 140
Dd	476 AGAGCAGTAGTGACACAGAGCCTGAACATTTGGTCTGTCTATAGAGATATCTTGAAATTCAT 535
Qy	141 ValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProaspAlaGlnLeuProSer 160



DT	22-OCT-2001	(first entry)	
XX	XX	Human polynucleotide SEQ ID NO 1232.	
DE	XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	XX	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	XX	leukemia; ss.	
OS	XX	Homo sapiens.	
XX	XX	WO200153312-A1.	
PN	XX	26-JUL-2001.	
XX	XX	26-DEC-2000; 2000WO-US34263.	
PF	XX	21-JAN-2000; 2000US-0488725.	
XX	XX	25-APR-2000; 2000US-0552317.	
PR	XX	09-JUL-2000; 2000US-0598042.	
PR	XX	19-JUL-2000; 2000US-0620312.	
PR	XX	03-AUG-2000; 2000US-0653450.	
PR	XX	14-SEP-2000; 2000US-0662191.	
PR	XX	19-OCT-2000; 2000US-0693036.	
PR	XX	29-NOV-2000; 2000US-0727344.	
XX	XX	(HYSB-) HYSEQ INC.	
XX	XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	XX	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX	XX	WPI; 2001-442253/47.	
DR	XX	P-PSDB; AAM39873.	
XX	XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	XX	such as central nervous system injuries -	
XX	XX	Claim 1; SEQ ID NO 1232; 10078pp; English.	
XX	XX	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	XX	the encoded polypeptides (AAM38642-AAM42213) with neurotropic,	
CC	XX	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	XX	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	XX	of the invention may be used to treat diseases of the peripheral nervous	
CC	XX	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	XX	localised neuropathies and central nervous system diseases, such as	
CC	XX	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	XX	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	XX	utilisation of the activities such as: Immune system suppression,	
CC	XX	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	XX	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	XX	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	XX	C.N.S disorders.	
CC	XX	Note: The sequence data for this patent did not form part of the printed	
CC	XX	specification.	
SQ	XX	Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;	
Alignment Scores:			
Pred. No.:	2,08e-227	Length:	1716
Score:	2386.00	Matches:	456
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.91%	Indels:	0
DB:	22	Gaps:	0
US-09-744-313A-1 (1-465) x AA159029 (1-1716)			
Qy	10	PheArgAsnThrGlnLysArgGlyGluSerPheGlyLysSerArgIleGlySerIle	29

Db	173	TTTCGGAAACACACAGAAAGGGAGAGATCATTTGGAAATCAGCAGATAGTAGCAAAAT	232
Qy	30	LyseGlyValPheLysSerThrThrMetGluGlyValaMetLeuProAsnTyrGlyVala	49
Db	233	AAAGGAGTATTCAAAAGTACCAACATGGAGGAGCTATGTTGCTTAATATATGTTAGCT	292
Qy	50	GluGlyGluAspPheLleGluGlyLleValValMetGluAspSerProVal	69
Db	293	GAAGGTGAAGATGATTTTATTTGAAGAGGTATTTGTTATTAAGAGATGATTTCCAGTG	352
Qy	70	GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro	89
Db	353	GAGCTGTGAGCACACCTAATATCTCCCGAAACCTTGCTGCTGGAATAATAGCATTTCCA	412
Qy	90	TyrValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro	109
Db	413	TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGAGAGAGAGAGAGAGAGATTCCT	472
Qy	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
Db	473	GTGTTTTGTATTGATGTTGAAGAAATGATAGAAGCAGTTCGACACAGAGCTGAACAT	532
Qy	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
Db	533	TGGTCTGCTATAGAAGATATCTTGAATTCATGTACTTGAATCAAACTAAACAGAAATTT	592
Qy	150	HisGlyValaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr	169
Db	593	CATGGTCAATTTCTGATGCCAGCTTCCTCTTAAGAGGATCATTTGGCCCCCAAAATAT	652
Qy	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrIleuGlnLysLeuLeuGlnHis	189
Db	653	GAATTTCTTAAGTCAAGAGGAGAGAGTTCCTCAAGAAATATCTACAGAACTTTCTGCAGCAT	712
Qy	190	ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
Db	713	CCAGAACTGAGTAATAGTCAACTCTGCGAGACTTTCTTTCCCTTAATGTTGGGGAAACA	772
Qy	210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro	229
Db	773	CAATTTCTTGAATAGTACTACAGAGATGTAATCTTTGGGAAATATTAATAATCTGTCTCT	832
Qy	230	GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
Db	833	GGAAACTATGAAAGAGAGAGAGTTCAGCATTTGGAACCTTTTATCATGATTTCAATAT	892
Qy	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
Db	893	TCITGTGAGTCTCCAAAGCTTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACT	952
Qy	270	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
Db	953	TCAGAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCCGTCTGAA	1012
Qy	290	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
Db	1013	AATACAGAGAGAGAGAGAGAGATTTTATGAGGAGTATGATGATGAGAGAGATC	1072
Qy	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis	329
Db	1073	TATGATTAACCTGATGATGATGAGAGAGGATTTTCCAGGTTCTGCTGCTTCATCAT	1132
Qy	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
Db	1133	CTCTTAATGGAACTCGAACTCTCTTTTAAACACCCCTGGAATGATGATGATGATGAT	1192
Qy	350	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
Db	1193	CTTCAGTGAACACTAGAACAGCTATTTTCAGAGCAGCCGTTTGGTCTCCTCACTAACACTT	1252
Qy	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
Db	1253	CTCAGAGATGCTATTTCTGTGAAACACTGAACTCGCTCTCTCTCCAGATAGACAAAA	1312



QY 270 SerGluAsnAsnLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnAlaGlu 289  
DB 1242 TCAGAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTGTGAA 1183  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
DB 1182 AATACAGAGAGAGACAAATCAGAAATATTTTATGAGGATGATGATGAGAGAGTC 1123  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
DB 1122 TATGATTACCTGATGATAGAGGGGTAGTTTCCAGGTTCCCTGAGCTGGCTCATCAT 1063  
QY 330 LeuLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1062 CTCCTTAATGGGAACCTGAAATCCTCTTTAAAAACACCTCGAATATGATATGATAT 1003  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369  
DB 1002 CTTCACTGTAACTAGAACAGCTATTTTCAAGAGCACCGTTTGTCTCACTCATACACTT 943  
QY 370 LeuArgAspAlaLeuPheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 942 CTCAGAGATGCTATATCTTGTGAAACACTGAACCTCGCTCTCTCAAGATAAGCAAAA 883  
QY 390 GlyAlaLysGlnThrPheGluMetMetAsnTyrIleProAspLeuValLysCys 409  
DB 882 GGAGCAAAACAGACTTTTGAAGAATGATGAATACATTCAGATCTGTAGTCAAGTGT 823  
QY 410 IleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
DB 822 ATTTGGTGAGAACCAAGATATGAACATCAGACTCTCTGTTGATGGCTTACAGCACCA 763  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
DB 762 GTACTACACAGAGAGCTGACTTATGTTTATTGGACATTTGTATACAGGAAGCTTTCCA 703  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
DB 702 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 655  
RESULT 4  
AAI60816/c  
ID AAI60816 standard; cDNA; 2176 BP.  
AC AAI60816;  
XX  
XX 22-OCT-2001 (first entry)  
DE Human polynucleotide SEQ ID NO 4805.  
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
\*PR

(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
WPI: 2001-442253/47.  
P-PSDB; AAM41660.  
Novel nucleic acids and polypeptides, useful for treating disorders  
such as central nervous system injuries -  
Claim 1; SEQ ID NO 4805; 10078pp; English.  
The invention relates to human nucleic acids (AAI57798-AAI61369) and  
the encoded polypeptides (AAM38642-AAM42213) with nontropic,  
immunosuppressant and cytostatic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
of the invention may be used to treat diseases of the peripheral nervous  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilisation of the activities such as: Immune system suppression,  
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
specification.  
SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
Alignment Scores:  
Pred. No.: 2.9e-227 Length: 2176  
Score: 2386.00 Matches: 456  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.91% Indels: 0  
DB: 22 Gaps: 0  
US-09-744-313A-1 (1-465) x AAI60816 (1-2176)  
QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
DB 2022 TTTCGGAACACACAGAAAGGGGAGAAATCATTTGGAAATCAGCAGATAGTAGCAAAAT 1963  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
DB 1962 AAAGGAGTATTCAAAAGTACCACCAATGGAGGAGCTATGTTGCTTAATTATGGTGTAGCT 1903  
QY 50 GluGlyGluAspAspPheIleGluGluGlyValValMetGluAspAspSerProVal 69  
DB 1902 GAAGTGGAAGATGATTTTATTGAAGAGGATTTGTTGTTATGAGAGATGATTTCTCCAGTG 1843  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaThrPylIleSerIlePro 89  
DB 1842 GAGGCTGTGACACACACTAATCTCTCCGAAACCTTCTGCTGATGAAATATAGCATTTCCA 1783  
QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109  
DB 1782 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGTTCCT 1723  
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
DB 1722 GTGTTTGTATGTATGTTGAAAGAAATCATAGAGAGCAGTTGGACAGAGCTTGAACAT 1663  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 1662 TGGTCTGTCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTAACAGAAATTT 1603  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169

Db 1602 CATGGTGCATTTCTGATGCCAGCTTCTCTTAGAGGATCATTTGGCCCAAAATTTAT 1543  
Qy 170 GluPheLeuLysSerIysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
Db 1542 GAATTCCTTAAGTCAAGAGGGAAGAGTTCCAAGATATCTACAGAAATCTTCTGCAGCAT 1483  
Qy 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 1482 CCAGAACTAGTAAATAGTCAACTCTCTGGCAGACTTTCTTTCCCTTAATGTGGGAAACA 1423  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValPro 229  
Db 1422 CAATTTCTTGATTAAGATATCTACAGATGTAATCTTTGGAAATATATAAATCTCTTCT 1363  
Qy 230 GlyLysLeuMetLysGlyGlnHisLeuGluProPheLeuMetAsnPheIleAsn 249  
Db 1362 GGAATACTAATGAAGAGAAAGTCAAGATTTGGAACCTTTTATCATGAATTTCAATTAAT 1303  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 1302 TCTTTGTGAGTCTCCAAAGCCCTAAACCAAGTAGACCAACTGACCAATCTTCAGCCCTACT 1243  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
Db 1242 TCAGAAACCAACAGAAAGCTTTTCAATGATCTGTTTAAATAATATCAACCGTCTGAA 1183  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 1182 AATACAGAGAAAGCAAAATCAGAAATTTATTTGAGGTGATGACTGTAGAGGAGTC 1123  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValPheGlnValPheGlnValProAspThrPheHis 329  
Db 1122 TATGATTACCTGATGATAGGACGGGTAGTTTCCAGGTCTCTGACTGGCTTCATCAT 1063  
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 1062 CTCTTAATGGGAATCGAACTCTTTTAAAAAACACCTCGAATGTATATCTGATTTACTAT 1003  
Qy 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
Db 1002 CTTGAGTGTAACTAGACAGCTATTTCAGGAGCACCGTTTGGTCTCACTCATACACTT 943  
Qy 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
Db 942 CTCAGAGATGCTATATTCTGTGAAACACACTGACCTCGCTCTCTCCAGATAAGCAAAA 883  
Qy 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
Db 882 GGAGCAAAACAGACTTTTGAAGAATGATGAATTTACATTCAGATCTGTAGTCAAGTGT 823  
Qy 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
Db 822 ATTGGTGAAAGAACCAAGATATGAAGCATCAGACTTCTGTTTGTATGGCTTACAGCAACA 763  
Qy 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
Db 762 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGTGATACAGGAATCTGTTTCCA 703  
Qy 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
Db 702 GAGCTCAATAGTACAAAGAGAGATTACCTCTGTGATCATCTTGGATG 655

## RESULT 5

ABQ79518

ID ABQ79518 standard; cDNA; 3145 BP.

XX

AC ABQ79518;

XX

DT 25-NOV-2002 (first entry)

XX

DE cDNA encoding a protein similar to human sorting nexin.

XX

KW Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;  
KW SSH 6; gene; human; nexin; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 128..2968

XX /\*tag= a

XX W0200262839-A2.

XX 15-AUG-2002.

XX 05-FEB-2002; 2002WO-EP01327.

XX 07-FEB-2001; 2001EP-0200439.

XX (UTMA-) UNIV MAASTRICHT.

XX Daemen MJAP, Cleutjens CBJM, Zaman GJR;

XX WPI; 2002-643400/69.

XX P-PSDB; ABB81193.

XX Use of a polynucleotide differentially expressed in ruptured and stable  
XX atherosclerotic plaques as a marker for atherosclerosis, useful in  
XX treating, diagnosing or preventing atherosclerosis -

XX Example 6; Page 35-37; 4pp; English.

XX The invention relates to the use of a polynucleotide differentially  
XX expressed in ruptured and stable atherosclerotic plaques as a marker  
XX for atherosclerosis, where the polynucleotides can be selected from the  
XX sequences shown in ABQ79517-19. The polynucleotides are useful as a  
XX marker of atherosclerosis, which may be used: (i) in the diagnosis,  
XX prevention and treatment of atherosclerosis; (ii) as serum/plasma markers  
XX to screen patients at risk for plaque instability to evaluate the effects  
XX of other treatments; (iii) in the preparation of vector molecules for the  
XX expression of the encoded protein in host cells; and (iv) in the  
XX identification of functional targets or analogues of the gene. The  
XX polynucleotides, the encoded proteins or antibodies against the proteins  
XX may be used to target other therapeutics to an unstable plaque.  
XX Modulation of the expression of the polynucleotide can increase plaque  
XX stability and therefore inhibit the progression of atherosclerotic  
XX cardiovascular disease. Modulators may be used to prepare pharmaceuticals  
XX for atherosclerotic disorders. The present sequence represents a cDNA  
XX encoding a protein similar to human sorting nexin.

XX Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;

## Alignment Scores:

Pred. No.:	1.21e-226	Length:	3145
Score:	2382.00	Matches:	454
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	99.56%	Mismatches:	0
Query Match:	97.74%	Indels:	0
DB:	24	Gaps:	0

US-09-744-313A-1 (1-465) x ABQ79518 (1-3145)

Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIysIle 29

Db 1598 TTTCGGAAACACAGAAAGGGGAGATCATTTTGGATCAGCAGATAGTAGCAAAATT 1657

Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49

Db 1658 AAAGGAGTATTCAGAGTACCACATGAGGGAGCTATGTTGCCATATTTAGTGTAGCT 1717

Qy 50 GluGlyGluAspPheIleGluGlyIleValMetGluAspSerProVal 69

Db 1718 GAAGGTGAAGATGATTTTATTCAAGAGAGTATTGTTGTAATGAAGATGATTCACAGTG 1777

Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89

Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;

Alignment Scores:			
Pred. No.:	8,846-195	Length:	1551
Score:	2057.50	Matches:	401
Percent Similarity:	87.94%	Conservative:	0
Best Local Similarity:	87.94%	Mismatches:	0
Query Match:	84.43%	Indels:	55
DB:	22	Gaps:	1
US-09-744-313A-1 (1-465) x AAI59030 (1-1551)			
QY	10	PheArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgIleGlySerLysIle	29
DB	173	TTTCGGAACACACAGAAAAGGGAGAAATCATTTGGAAATCAGAGAAATAGTAGCAAAAT	232
QY	30	LysGlyValPheLysSerThrMetGluGlyValMetLeuProAsnTyrGlyValAla	49
DB	233	AAAGAGATATTCAAAGTACCAATATGGAGGAGCTATGTTGCTTAATTTATGCTAGCT	292
QY	50	GluGlyGluAspPheIleGluGlyValValMetGluAspSerProVal	59
DB	293	GAAGGTGAATGATTTTATGAAGAAGGTATTTGTTATGGAAGATGATTTCTCCAGTG	352
QY	70	GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro	89
DB	353	GAGGCTGTGAGCACACCTAATATCTCCCGAAACCTTCTGCATGGAAATTTAGCATTC	412
QY	90	TyrValAspPheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro	109
DB	413	TATGTAGCTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAGAAAAGAAATTCCT	472
QY	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
DB	473	GTGTTTGTATGATTTGAAGAAATGATAGAGACAGCTTGGACACAGAGCTGAACAT	532
QY	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
DB	533	TGGTCTGTCTATAGATATCTTGAATTTCTATGTACTTGAATCAAAATCAACAGATTT	592
QY	150	HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr	169
DB	593	CATGGTGCATTTCTGTAGTCCAGCTTCTTCTTAAGAGATCATTTGCCCCCAAAATAT	652
QY	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis	189
DB	653	GAATTCCTTAAGTCAAGAGGGAAGAGTTCCAGAAATATCTACAGAAATCTTCAGCAT	712
QY	190	ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
DB	713	CCAGAACTGAGTAAATAGTCAACTCTTGGCAGACTTTCTTCCCTAATATGTTGGGAA	772
QY	210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValPro	229
DB	773	CAATTTCTTGAATAGTACTACAGATGATAATCTTGGGAAATATAAATCTGTTCT	832
QY	230	GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
DB	833	GGAAACTAATGAAGAGAGAGGTTCAGATTTTGGACCTTTTATCAATGAATTCATTAAT	892
QY	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
DB	893	TCTTGTGAGTCTCCAAAGCTAAACCAAGTAGACAGCAACTGACCAATCTCAGCCCTACT	952
QY	270	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
DB	953	TCAGAAAAACAACAAGAGCTTTTCAATGATCTCTTAAATAATATGCAACCGTCTGAA	1012
QY	290	AsnThrGluArgGlyGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
DB	1013	AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGGAGGTGATGACGTAGAGGATC	1072
QY	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisHis	329
DB	1073	TATGATTAATCTGATGATGTA	1093

QY	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
DB	1093	-----	1093
QY	350	LeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
DB	1093	-----	1093
QY	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
DB	1094	-----GATGCTATATCTGTGAAACACATGAACCTCGCTCTCTCCAGATAGCAAAA	1147
QY	390	GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys	409
DB	1148	GGACCAAAACAGACTTTTGAAGAAATGATGAATACATTCACATCTGTTAGTCAAGTGT	1207
QY	410	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro	429
DB	1208	ATTGCTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGTGATGGCTTTACAGCA	1267
QY	430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449
DB	1268	GTACTCAACAGCAGCTGACTTATGTTTATTTGGACATTTGATACAGGAACTGTTTCCA	1327
QY	450	GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet	465
DB	1328	GAGTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG	1375
RESULT 7			
ID	AAZ17379	standard; cDNA; 779 BP.	
XX	AAZ17379;		
XX	12-OCT-1999 (first entry)		
DE	Human gene expression product cDNA sequence SEQ ID NO:4852.		
XX	Human; gene; gene expression product; diagnosis; therapy; probe;		
KW	detection; mapping; tissue typing; profiling; forensic; cancer;		
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.		
XX	Homo sapiens.		
XX	WO9938972-A2.		
XX	05-AUG-1999.		
XX	28-JAN-1999; 99WO-US01619.		
XX	03-APR-1998; 98US-0080666.		
PR	28-JAN-1998; 98US-0072910.		
PR	24-FEB-1998; 98US-0075954.		
PR	31-MAR-1998; 98US-0080114.		
PR	03-APR-1998; 98US-0080515.		
XX	(CHIR ) CHIRON CORP.		
PA	(HYSE-) HYSEQ INC.		
XX	Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;		
PI	Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;		
PI	Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;		
PI	Lamson G, Leskowitz D, Pot D, Randazzo F, Reinhard C;		
PI	Stache-Crain B, Sudduth-Klinger J, Williams LT;		
XX	WPI; 1999-494092/41.		
XX	Novel human genes and their expression products which are		
PT	differentially expressed in different cell types		
XX	Claim 1; Page 2307; 2479pp; English.		
PS			
XX			

The present invention describes a library of human polynucleotides comprising the sequences given in AA121532 to AA217779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA121532 to AA217779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purposes, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

Sequence 779 BP: 248 A; 143 C; 153 G; 232 T; 3 other; XX \$0

Alignment Scores:		
Pred. No.:	2.44e-95	779
Score:	1054.00	205
Percent Similarity:	99.04%	Conservative: 2
Best Local Similarity:	98.09%	Mismatches: 1
Query Match:	43.25%	Indels: 1
DB:	20	Gaps: 0

US-09-744-313A-1 (1-465) x AAZ17379 (1-779)

257	QY	LysProSerArgpProGlu-LeuThrIleLeuSerProThrSerGluAsnAenLysLe	276
:		.....	
61	Db	AGACCAAGTAGACCAGAAACTGACCATTCCTCAGTCTACTTCAGAAACAACAAGAGCT	120
276	QY	uPheAsnAspLeuPheLysAenAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAs	296
121	Db	TTTCAATGATCTGTTTAAAAATAATGCAAAACGGTCTGAAATAATACAGAGAGAAAGCAAAA	180
296	QY	nGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVa	316
181	Db	TCAGAAATATTATTTATGGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGGATGTATGT	240
316	QY	IGLYArgValValPheGlnValProAspTyrLeuHisIleuLeuMetGlyThrArgIl	336
241	Db	AGGACGGGTAGTGTTCAGGTTCCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAAT	300
336	QY	eLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGl	356
301	Db	CCTCTTTAAAAACACCCCTGGAAATGTATCTGATTACTATCTTCAGTGTAAACTAGAAC	360
356	QY	nLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCy	376
361	Db	GCTATTTTCAGGAGCACCGTTGGTCTCACTCATAACACTTCTCAGAGATGCTATATTCTG	420
376	QY	sGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGl	396
421	Db	TGAAACACACTGAACCTCGCTCTCTCCAAGATAAGCAAAAAGGAGCAAAACAGACTTTTGA	480
396	QY	uGluMetMetAsnTyrIleProAspLeuValLysCysIleGlyGluGluThrLysTy	416
481	Db	AGAAATGATGTAATTCATTCAGATCTGTATGTCAGGTGTATTTGGTGAAGAACCAAGTA	540
416	QY	rdLusertIleArgLeuLeuPheAspGlyLeuGlnProValLeuAsnLysGlnLeuTh	436
541	Db	TGAAAGATCATGACTTCTGTTTGTATGGCTTACAGCAACGACTACTCAACAACGACGTGAC	600
436	QY	rtYrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLys	456
601	Db	TTATGTTTATTGGACATTTGTGATCAGGAACCTGTTTNCAGAGCTCAATAAGTCTCAAAA	660

Qy 456 sGIValThrSerValThrSerTrp 464  
db 661 GGAAGTTACCTCTGTGACATCTTG 685

## RESULT 8

ABS77486  
ID ABS77486 standard: cDNA: 725 BP.

XX  
XX  
XX

AC  
XX  
ABS / 1406;DT 12-DEC-2002 (first entry)  
XX

DE Frog embryonic gene sequence Q9925894.

Frog; ss; embryonic development; de-

KW Microarray; cell  
XX

OS Xenopus laevis.  
XX

PN US2002081610-A

27-JUN-2002.

XX  
PF 23-JUL-2001; 2001US-0910943.

XX  
PR 21-JUL-2000: 2000US-219658[illegible]XX  
FA (UIRQ) UNIV ROCHESTER

PI	Hemmati-Brivanlou A, Altmann CK;
XX	WFI; 2002-G26534/67.
DR	
XX	
XX	Nucleic acid array containing Xenopus embryonic nucleic acids is useful
PT	to identify genes involved in embryonic development, to identify
PT	different types of embryonic cells, and to diagnose developmental
PT	disorders -

XX  
ps  
Claim 1: page 756: English:

XX The invention relates to a nucleic acid array, where each coordinate  
CC contains a single nucleic acid species having one of 770 nucleotide  
CC sequences (appearing as ABS76747-ABS7516) a of a Xenopus embryonic gene  
CC product, or its complement or hybridisable fragment of not less than  
CC 20 contiguous nucleotides of one of those sequences. Also, included are  
CC detecting differential expression of embryonic genes, comprising:  
CC (a) contacting a nucleic acid array comprising genes expressed in  
CC embryonic but not mature cells with nucleic acids from sample and  
CC control cells; and (b) detecting differential hybridisation of nucleic  
CC acids from the sample cells relative to the control cells; and detecting  
CC defects in development, comprising: (a) contacting nucleic acids from  
CC test cells undergoing development with a nucleic acid array of gene  
CC products known to play a fundamental role in the development process; and  
CC (b) detecting a difference in expression of a fundamental gene in the  
CC sample cells relative to a standard. The invention is useful to identify  
CC genes involved in embryonic development and related processes such as  
CC cell differentiation. This would be useful for diagnosing developmental  
CC disorders and for identifying different types of embryonic cells.  
CC The present sequence is one of the 770 Xenopus embryonic cDNA sequences.

XX  
90  
sequence 725 bp: 229 A; 135 C; 140 G; 218 T; 3 other;

**Alignment Scores:**

Pred. No.:	3 62e-88	Length:	725
Score:	91.50	Matches:	193
Percent Similarity:	85.83%	Conservative:	13
Best Local Similarity:	80.42%	Mismatches:	29
Query Match:	40.27%	Indels:	5
	24	Gaps:	1
			DB:

US-09-744-313A-1 (1-465) x ABS77486 (1-725)



QY 187 LeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAaspPheLeuSerProAsnGly 206  
Db 20 TTGCAGGATCCCATCGAATTCGAATTCGTCGACCCACGCG-----TCGGT 64  
QY 207 GlyGluThrGlnPheLeuAspLysLeuProAspValAsnLeuGlyLysLeuLys 226  
Db 65 TTGGAAACTCAGTTTCATGACAAATCCCTACAGATGTAATCTGGAAATGATCAAG 124  
QY 227 SerValProGlyLysLeuMetLysGlyGlnHisLeuGluProPheLeuMetAsn 246  
Db 125 TCGATTCCTGGAAACTCATTAAGAAAGGCGACATCTGGAGCTGTTTATATGAT 184  
QY 247 PheLeuAsnSerCysGluSerProLysProLysProSerArgProGluLeuThrLeu 266  
Db 185 TTCATTAACATCATGTGAATCACCACCAACCAACCAAGTAGGCTGAACTGACTATTCTA 244  
QY 267 SerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsn 286  
Db 245 AGCCCGACTTTCAGAAACCAATAAAGCTTTTAAATGATCTGTACAGAAATATGCTAAT 304  
QY 287 ArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrVal 306  
Db 305 CGCTCTGAAATACAGAAAGGAGCATACCAAGATTAATTTATGGAAGTGTACCGTA 364  
QY 307 GluGlyValTyrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrp 326  
Db 365 GAGGGTGTATGACTATTATATGATATAGGACGCTGTGTTTTCACATTCCTGATGG 424  
QY 327 LeuHisLeuLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThr 346  
Db 425 TTTTCATCATCTTTGATGGTGGACGAATCTCTTTAAACATACATCCCTCGAGACTTACACA 484  
QY 347 AspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeu 366  
Db 485 AATAGCTATTAACTATAATTAATAGACGCTCTTTCAAGAGATCGCTGTTTCTTTG 544  
QY 367 IleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAsp 386  
Db 545 ATTACTCTGCTTAGAGATGCTATATTGTAACCGCTGAACACGATCATACATCG 604  
QY 387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
Db 605 AAGCAGCAAGGGGCAAGCTTACATTTGAAGAAATGATGCGCTATATTCAGATTGTGAT 664  
QY 407 ValLysCysLeuGlyGluGluThrLysTyrGluSerIleArgLeuPheAspGlyLeu 426  
Db 665 GGTAATGATGTTGCTGATGAAGCTAAATATGAAGGCATGCGACTTCTGTTTGGTGACTG 724

RESULT 9

ID ABX50611 standard; cDNA; 402 BP.

AC ABX50611;

XX 25-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #540.

DE Bovine EST associated with lactation/muscle/fat deposition #540.

XX Muscle; ss; EST; expressed sequence tag; lactation; LMFD;

KW muscle deposition; fat deposition; genome mapping; gene identification;

KW gene analysis; cattle breeding.

XX Bos Taurus.

OS US2002137160-A1.

XX 26-SEP-2002.

XX 26-OCT-2001; 2001US-0983965.

XX 17-DEC-1998; 98US-113678P.

PR 15-DEC-1999; 99US-0455231.

XX

PA (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-102386/09.  
DR Purified nucleic acid molecules, useful for genome mapping, gene  
XX identification and analysis, cattle breeding or preparation of  
XX constructs for cattle gene expression and genetically improved cattle -  
PT Claim 2; SEQ ID No 540; 39pp; English.  
PT The invention relates to a purified nucleic acid molecule associated with  
XX lactation or muscle and fat deposition (designated LMFD), derived  
XX from cattle, and the LMFD nucleic acid can specifically hybridise to a  
XX second nucleic acid molecule comprising any of 5912 nucleotide  
XX sequences, appearing as ABX50072-ABX59983, or complements of them.  
XX Also included are; (1) a transformed cell having a nucleic acid  
XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
XX translated sequence that functions in the cell to cause termination of  
XX transcription and addition of polyadenylated ribonucleotides to a 3' end  
XX of the mRNA molecule; and (2) determining a level or pattern of a  
XX molecule in a bovine cell or tissue comprising: (a) incubating a marker  
XX nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
XX complement or fragment) with a complementary nucleic acid molecule  
XX obtained from the bovine cell or tissue, where hybridisation between the  
XX marker nucleic acid and the complementary nucleic acid permits the  
XX detection of the molecule; and (b) detecting the level or pattern of the  
XX complementary nucleic acid, where the detection of the complementary  
XX nucleic acid is predictive of the level or pattern of the molecule.  
XX The LMFD nucleic acid is used for determining a level or pattern  
XX of a molecule in a bovine cell or tissue. It is useful for genome  
XX mapping, gene identification and analysis, cattle breeding, preparation  
XX of constructs for use in cattle gene expression, or for genetically  
XX improving cattle. The present sequence is one of the 5912 bovine  
XX LMFD EST (expressed sequence tag) nucleic acids.  
XX Note: The present sequence was not shown in the specification but  
XX was obtained in electronic format from the USPTO web site:  
XX seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX Sequence 402 BP; 141 A; 71 C; 76 G; 113 T; 1 other;

Alignment Scores:  
Pred. No.: 1.01e-58 Length: 402  
Score: 682.00 Matches: 127  
Percent Similarity: 99.25% Conservative: 5  
Best Local Similarity: 95.49% Mismatches: 1  
Query Match: 27.99% Indels: 0  
DB: 25 Gaps: 0

US-09-744-313A-1 (1-465) x ABX50611 (1-402)

QY 205 AsnGlyGlyGluThrGlnPheLeuAspLysLysLeuProAspValAsnLeuGlyLysLeu 224  
Db 3 AATGGTGGGAAACACAGATCTTCTTGATAAGATACTACCAGACGTAATCTTGGAAAT 62  
QY 225 IleLysSerValProGlyLysLeuMetLysGlyGlnHisLeuGluProPheLeu 244  
Db 63 ATAAATCTGTTCTCTGAAAACCTAATGAAGAGAAAGTCAACATTTTGGACCTTCATC 122  
QY 245 MetAsnPheLeuAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr 264  
Db 123 ATGAATTCATTAATTCCTTGTGAATCTCTCAAGGCTTAACCGATTAACCGAATCTGACC 182  
QY 265 IleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsn 284  
Db 183 ATTCTACCCCTTACTCTAGAGATTAATAGAGCTTTTATGATCTGTATAGATAAT 242  
QY 285 AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMet 304



Db 243 GCRAACCGTCTGAGAAATACAGAAAGGCAAAATCAGAAATATTTATGGAATGATG 302  
 Qy 305 ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPro 324  
 Db 303 ACTGTAAAGAGTCTATGATTTACCTGATGTAAGTANGACAGTGGTTTCCAGTTCT 362  
 Qy 325 AspTyrLeuHisLeuLeuMetGlyThrArgGileLeu 337  
 Db 363 GACTGGCTTCATCATCTCTTAATGGAAGTCAATCCTC 401  
 RESULT 10  
 AAC02755  
 ID AAC02755 standard; cDNA; 358 BP.  
 AC AAC02755;  
 XX  
 DT 06-OCT-2000 (first entry)  
 DE Human secreted protein 5' EST, SEQ ID NO: 2753.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000BP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 DR P-PSDB; AAG02749.  
 XX  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 -PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 XX Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX  
 SQ Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 5.15e-54 Length: 358  
 Score: 634.00 Matches: 119  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.02% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-744-313A-1 (1-465) x AAC02755 (1-358)

Qy 216 LeuProAspValAsnLeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGln 235  
 Db 2 CTACCAGATCTAAATCTTGGAAAAATATAAAATCTGTTCTCTGAAAACTAATGAAAGAG 61  
 Qy 236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLys 255  
 Db 62 AAAGGTGAGCATTTGGAACTTTTATCATGAATTTTCAATTAATCTTTGTGAGTCTCAAAG 121  
 Qy 256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLys 275  
 Db 122 CTTAAACCAAGTAGACAGCACTGACCATCTCTCAGCCCTACTTCAGAAACACACAGAG 181  
 Qy 276 LeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
 Db 182 CTTTTCATGATCTGTTTAAAAATAATGCAACCGTGTGAAATAACAGAGAGAAAGCAA 241  
 Qy 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315  
 Db 242 AATCAGAATTAATTTATGGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGTAT 301  
 Qy 316 ValGlyArgValValPheGlnValProAspTyrLeuHisLeuLeuMetGlyThr 334  
 Db 302 GTAGGACGGGTAGTTTTCAGGTTCCTGCTGCTTCATCATCTCTTAATGGGAACC 358  
 RESULT 11  
 ABX50426  
 ID ABX50426 standard; cDNA; 305 BP.  
 XX  
 AC ABX50426;  
 XX  
 DT 25-FEB-2003 (first entry)  
 DE Bovine EST associated with lactation/muscle/fat deposition #355.  
 XX  
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
 KW muscle deposition; fat deposition; genome mapping; gene identification;  
 KW gene analysis; cattle breeding.  
 XX  
 OS Bos Taurus.  
 XX  
 PN US2002137160-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 26-OCT-2001; 2001US-0983965.  
 XX  
 PR 17-DEC-1998; 98US-113678P.  
 PR 15-DEC-1999; 99US-0465231.  
 XX  
 PA (BYAT//) BYATT J C.  
 PA (MATH//) MATHIALAGAN N.  
 PA (TAON//) TAO N.  
 PA (WARR//) WARREN W C.  
 XX  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 XX  
 WPI; 2003-102386/09.  
 XX  
 PT Purified nucleic acid molecules, useful for genome mapping, gene  
 PT identification and analysis, cattle breeding or preparation of  
 PT constructs for cattle gene expression and genetically improved cattle -  
 XX  
 PS Claim 2; SEQ ID No 355; 38pp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMFD), derived to a  
 CC from cattle, and the LMFD nucleic acid can specifically hybridize to a  
 CC second nucleic acid molecule comprising any of 5912 nucleotide  
 CC sequences, appearing as ABX50072-ABX55983, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end

of the mRNA molecule; and (2) determining a level or pattern of a molecule in a bovine cell or tissue comprising: (a) incubating a marker nucleic acid (comprising any of the 5912 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridisation between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule. The LMFD nucleic acid is used for determining a level or pattern of a molecule in a bovine cell or tissue. It is useful for genome mapping, gene identification and analysis, cattle breeding, preparation of constructs for use in cattle gene expression, or for genetically improving cattle. The present sequence is one of the 5912 bovine LMFD EST (expressed sequence tag) nucleic acids.

Note: The present sequence was not shown in the specification but was obtained in electronic format from the USPTO web site: [seqdata.uspto.gov/sequence.html?DocID=20020137160](http://seqdata.uspto.gov/sequence.html?DocID=20020137160).

**Homo sapiens.**

WO200216388-A1.

28-FEB-2002.

17-JAN-2001; 2001WO-US01394.

18-AUG-2000; 2000US-226278P.

05-JAN-2001; 2001US-259805P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;  
Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Piscella M;  
Ni J;

WPI; 2002-292053/  
P-PSDB; AAU91368.

Isolated nucleic acids encoding 21 secreted proteins useful for the diagnosis and treatment of e.g. cancer, HIV infection, stroke and rheumatoid arthritis -

Claim 1; Page 451; 54pp; English.

The present invention relates to the isolation of novel human secreted proteins, and the polynucleotide sequences encoding them. The secreted proteins are useful to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The secreted proteins are also useful in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the secreted proteins can also be used in alleviating symptoms associated with disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which can be diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. cancer, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Parkinson's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The amino acid sequences of the secreted proteins are as follows:

sequence 1900 BP: 657 A: 328 C: 400 G: 512 T: 3 other;

RESULT 13  
AAI61094

RESULT 13  
AAI61094

QY	397	GluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr	416
DB	1416	AAGCTGCTGAAACAACTCCAGATATGCTTCAGACCTTGTTCAGACCAAAATGCCCGC	1475
QY	417	GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr	436
DB	1476	CACGGTATAATAAAATATTCAATCACTGCAAGAAACAAGAGCAACAAGCATCTGTTA	1535
QY	437	TyrValLeuLeuAspIleValleGlnGlnLeuLeuPheProGluLeu	451
DB	1536	TATGCGCTGATGAAGTCTGCTGAATTGAATGCTGCTGCTGAGCTG	1580
RESULT 14			
AAI59308			
ID	AAI59308	standard; cDNA; 2181 BP.	
XX	XX	AAI59308;	
XX	XX	22-OCT-2001 (first entry)	
DT	XX	Human polynucleotide SEQ ID NO 1511.	
XX	XX	Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
XX	XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	KW	leukaemia; ss.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX	WO200153312-A1.	
FN	XX	26-JUL-2001.	
PD	XX	26-DEC-2000; 2000WO-US34263.	
PP	XX	21-JAN-2000; 2000US-0498725.	
XX	XX	25-APR-2000; 2000US-0552317.	
PR	PR	09-JUL-2000; 2000US-0598042.	
PR	PR	19-JUL-2000; 2000US-0620312.	
PR	PR	03-AUG-2000; 2000US-0653450.	
PR	PR	14-SEP-2000; 2000US-0662191.	
PR	PR	19-OCT-2000; 2000US-0693036.	
PR	PR	29-NOV-2000; 2000US-0727344.	
XX	XX	(HYSE-) HYSEQ INC.	
PA	XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;	
PI	PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
PI	PI	WPI; 2001-442253/47.	
DR	DR	P-PSDB; AAM40152.	
XX	XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	PT	such as central nervous system injuries -	
PT	PT	Claim 1; SEQ ID NO 1511; 10078pp; English.	
PS	XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and	
XX	XX	the encoded polypeptides (AAM38642-AAM42213) with nontropic,	
CC	CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	CC	localised neuropathies and central nervous system diseases, such as	
CC	CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	CC	utilisation of the activities such as: Immune system suppression,	
CC	CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	

US-09-744-313A-1 (1-465) x AA161094 (1-2133)
56 IleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal 72
531 TTGAAGGATGAATAATCTTAATAGAGAAAGAACGACAGACCTTCAGCTGCACATGGCA 590
73 SerThrProAsnThrProArgAsnLeuAlaIlePheIleSerIleProTyrValAsp 92
591 AGAACGGATTGGTGGTGTGAAGAACCTTGGCATGTGGAAGAGCTCATCACAGTGGAGAG 650
93 PhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCys 112
651 GTTACAGAA-----GAGATGGTGGAGCAATGCCATGTACTTT 689
113 IleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyrSerVal 132
690 GTCATGGTAAAGCCTACAAGAA-----GTTGGAGGAGTTGAAACTAGAACTGACGCGTC 743
133 TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAla 152
744 CCCAAAGGCTCAGCGAGTTTCATTAATTTACACCGGAAACTCAGTGAAGTGGTTCCTTCT 803
153 PheProAspAlaGluLeuProSerLysArgIleIleGlyProLys-----AsnTyrGlu 170
804 TTAATAAAGACAGTGCCTTCTCTTAGCAAGCTGCCCTTCAATCATATAGATCACACG 863
171 PheLeuLysSerLysArgGluGluPheGluGluTyrLeuGluLysLeuGluHisPro 190
864 TTTATGGAAGACTTGGAGTAATCAATTAATAAGTTTTACAGAACTGCTTTCAGATCAA 923
191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210
924 AGACTGTGTGACAGTGAAGCACTTTATGCTCTCTTGAGCCCTTCTCTGACTACCTCAAG 983
211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGly 230
984 GTTATCGAC-----GTGCAGGGA 1001
231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250
1002 -----AAAAAATCTTTTCATTAATCTCTCATTTTG----- 1034
251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270
1034 ----- 1034
271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290
1035 -----GAAAGACTTCTCGCGACTTCTTC-----TCCACACGAGGAGGAG 1076
291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310
1077 ACAGAGGAGACAGTGACCTGTCAGATTATGTGAT-----GATGTGATGGGAGGAAA 1130
311 AspTyrLeu-----MetTyrValGlyArgValValPheGlnValPro----- 324
1131 GACGCTTGGCTGAACCATGTTTCATGTTGATTGGGAG-----ATTTTTGAACCTCGAGGA 1187
325 -----AspTyrLeuHisLeuLeuMetGly---ThrArgIleLeuPheLysAsnThr 341
1188 ATGTTTAAATGGGTGAGAGAAACATTAATTTGCCCTCGTTTCAGGTCACTTTTGAAGAAC 1247
342 LeuGluMetTyrThrAspTyrTyrLeuGlnCysLeuLeuGluGlnLeuPheGlnHis 361
1248 ATC-----AACAAACAAATCCCGGACACAGTCAGCTGGGATTTTCAGTGAAGCAA 1295
362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePhe-----Cys 376
1296 ATGTTGGTTTACTACATCAATATTTTCGGGATGCTTTTGGCCAAATGGGAGTTGGCA 1355
377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396
1356 CCACCGACCAATCAGAACCAAGACCAAGTCAAGAAACAAACAAACAGAGACACAGCAA 1415



The proteins are useful for identifying compounds that modulate their activity and production. The sequences of the invention are useful for treating diseases such as hyperproliferative disorders (e.g. cancer), immune deficiency disorders (e.g. severe combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis), infectious disorders (e.g. hepatitis), allergic conditions (e.g. asthma), neurodegenerative disorders (e.g. Alzheimer's disease), liver fibrosis, coagulation disorders (e.g. haemophilia), and tumours. The polynucleotide sequences of the invention are also useful in gene therapy. AK35610-AK3232 represent the cDNA sequences of the invention that encode for novel human secreted proteins.

Sequence 2494 BP: 845 A; 445 C; 547 G; 657 T; 0 other;

Alignment Scores:	
Pred. No.:	1,398-14
Score:	249.50
Percent Similarity:	42.75%
Best Local Similarity:	24.15%
Query Match:	10.24%
DB:	24
	Gaps: 14
	Indels: 83
	Mismatches: 155
	Conservative:
	77
	Matches: 100
	Length: 2494

US-09-744-313A-1 (1-465) X ABK35847 (1-2494)

Qy	56	IleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal	72
Db	908	TTGAAGCATGAAATAATCTCTAATAGAGAAAGACGACAGACTTCAGTGCACATGGCA	967
Qy	73	SerThrProAsnThrProArgAsnLeuAlaAlaTlPylsIleSerIleProTyrValAsp	92
Db	968	AGAACGATGTGGTGTGTGAAACCTTGGCATGTGGAAAGCCTCCATCACCACGTGAGAG	1027
Qy	93	PhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCys	112
Db	1028	GTTCACAGAA-----GAGAAATGGTGAGCAATGGCCATGTTACTTT	1066
Qy	113	IleAspValGluArgAsnAspArgArgAlaValGlyHisGluProGluHisTlPylSerVal	132
Db	1067	GTTCATGTTAAGCCTACACAGAA-----GTTGGAGGAGTTGAAACTAAGAACTGCACGGTC	1120
Qy	133	TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLysLeuThrGluPheHisGlyAla	152
Db	1121	CCCAGAAGGCTCAGCGAGTTTCAGAAATTACCGGAAACTCAGTGAAGTCGCCCTCT	1180
Qy	153	PheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu-----	170
Db	1181	TTAAAAAAGTCAGTGGCTTCTCTTAGCAGAGCTGCCTTTCAAATCTAGATCAAAAG	1240
Qy	171	PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisPro	190
Db	1241	TTTATGGAAAGTCGAAGATCAATTAATAAGTTTTTACAGAATCTGCTTTTCAGATGAA	1300
Qy	191	GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln	210
Db	1301	AGACTGTGTACAGAGTGAAGCACTTTATGGCTTCTTTGAGCCCTTCTCCTGACTACCTCAAG	1360
Qy	211	PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGly	230
Db	1361	GTATATCGAC-----GTGACAGGG	1378
Qy	231	LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer	250
Db	1379	-----AAAAAAATCTTTTTCATTATCTCTCATTTTGT-----	1411
Qy	251	CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer	270
Db	1411	-----	1411
Qy	271	GluAsnAsnLysLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn	290
Db	1412	-----GAAAGACTTCTCGGAGACTTCTTC-----TCCACACGAGGAGGAG	1453

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:43:52 ; Search time 2927 Seconds  
(without alignments)  
3861.148 Million cell updates/sec

Title: US-09-744-313a-1  
Perfect score: 2437  
Sequence: 1 MYLHLCIFRNTQKRGESF.....ELFPELNKVKQKVTSTSM 465

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO/US09744313/runat 30012004\_113535\_4196/app\_query.fasta\_1.647  
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-USER=US09744313 @CNG 1 1 2810 @runat 30012004\_113535\_4196 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_etc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estcom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_man.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_pig.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2347	96.3	3064	11	AK049371	Mus muscu
2	1949	80.0	2634	11	AK082389	Mus muscu
3	1517	62.2	917	13	BX372281	BX372281
4	1404.5	57.6	905	14	CB182722	CB182722
5	1348.5	55.3	1043	9	AL574298	AL574298
6	1345	55.2	994	13	BUI18410	BUI18410
7	1271	52.2	818	10	BG191447	BG191447
8	1266	51.9	1096	12	BQ053127	BQ053127
9	1230.5	50.1	921	13	BQ230190	BQ230190
10	1220	50.1	903	9	AL578668	AL578668
11	1190.5	48.9	915	13	BQ243572	BQ243572
12	1190	48.8	1105	10	BQ256631	BQ256631
13	1172	48.1	1093	10	BQ293966	BQ293966
14	1162	47.7	754	13	BQ273460	BQ273460
15	1151.5	47.3	1076	10	BQ874550	BQ874550
16	1148	47.1	664	14	CB136167	CB136167
17	1127	46.2	1096	14	CD500048	CD500048
18	1117	45.8	656	10	BQ386253	BQ386253
19	1111	45.6	722	9	AJ393454	AJ393454
20	1084	44.5	625	10	BQ536529	BQ536529
21	1083	44.4	643	14	CB114674	CB114674
22	1069.5	43.9	638	10	AW957704	AW957704
23	1059.5	43.9	726	13	BX303054	BX303054
24	1064	43.7	952	13	BQ171610	BQ171610
25	1059.5	43.5	810	10	BG191057	BG191057
26	1059	43.5	608	12	BQ722250	BQ722250
27	1059	43.5	752	13	BQ282728	BQ282728
28	1056	43.3	608	12	BQ29345	BQ29345
29	1053	43.2	616	14	CB054303	CB054303
30	1045	42.9	705	13	BX17013	BX17013
31	1045	42.9	936	10	BQ026479	BQ026479
32	1044	42.8	604	14	CB130920	CB130920
33	1042	42.8	802	14	CA431124	CA431124
34	1038	42.6	606	12	BQ699448	BQ699448
35	1025.5	42.1	861	10	BG214306	BG214306
36	1021	41.9	817	10	BQ692236	BQ692236
37	1020	41.9	897	10	BQ180026	BQ180026
38	995	40.8	828	13	BUI13631	BUI13631
39	988.5	40.6	936	13	BQ36152	BQ36152
40	979.5	40.2	800	10	BQ581597	BQ581597
41	976	40.0	656	9	AL957064	AL957064
42	972	39.9	675	10	BG410145	BG410145
43	968	39.7	558	14	CB160118	CB160118
44	963	39.5	923	9	AL561693	AL561693
45	962	39.5	639	13	BQ333500	BQ333500

ALIGNMENTS

RESULT 1  
AK049371  
LOCUS  
DEFINITION  
Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
clone: C330035N22 product: HYPOPHYSICAL 102.9 KDA PROTEIN (SORTING  
NEXTIN 14) homolog [Homo sapiens], full insert sequence.  
ACCESSION  
AK049371.1 GI:26340105  
VERSION  
AK049371.1  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1	Carrincci, P. and Hayaashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636	COMMENT	Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/.
REFERENCE	2	Carrincci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayaashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159	FEATURES	Location/Qualifiers 1. 3064 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM:DB:C330035N22" /db_xref="taxon:10090" /clone="C330035N22" /cell_type="ES cells" /clone_lib="RIKEN full-length enriched mouse cDNA library" 124. 2937 /note="unnamed protein product; HYPOHETICAL 102.9 KDA PROTEIN (SORTING NEXIN 14) homolog [Homo sapiens] (SPTRIQ9BSD1, evidence: PASTY, 93.3%ID, 99.6%length, match=2781) putative"
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Ozawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayaashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861	CDS	protein_id="BAC33716.1" /db_xref="GI:126340106" translations="MGSWRTICGRLLKQRLDVGREICRQYPLFCFLLCGLSVASILL LVKLVHLMFVSGVAVVTFYGLSDPLNFIFFTKYKQGLQBLPQGHSCA CGVKCKRHRPSILLNYPWLDLVSSKVDASLSEVLELVLENVYFWYRWDTDE SFVDELRTLRPFASVLRIRHKVDIPSIITKLLKAKMKHIEVVKRQKVNTEYL QQALBEYGPBLHVALNRDELQYLRKLTLLPPLPKADTCSLTLLEILLSG SVLPSLDFLADPTVHLLIIIDDSPPPEKAPSPVPLQKFAEPKPKKPSVLK LEKIQREQQLLFRFMFLKQGAHVLFQCLTFEFDNRLIRLPELNDMLSLHEE LOKITYCLDSEIDKIRFPFVIEIRIAGFYIDVVKLOTWRCLEFAYEHLVLSL ENVTMPKCHSDVYFCLLAGSPTEKNSFSTOKRGSGFISGSKIKGWFKST TMEGAVLVKNGAEGDDFTEBGIIVMDDSPVEAVSTPNTENLAWKISIPVDFF EDSSERKKERKIPVFCIDVRRNDRRAVGHPEHWSVRYRLFFVLESKLEFPHGT FPDQLPSRLIIIGPKYFELKSRFEQKLQVLPSELSKPLDFLSPNGEFTQ FLDLKPLNGLIKIISVPGKLMKKEQHLPEFIMFINSCESPKSPFSPHTILSP TSENKLENDLPKNANRAENTERKQNYFMEVMTDGVYDLYLVYGRVQVDPDM LHLMLTRILFKNTLENYDYVLOCKLEQLQFHEHVLVSLITLRDAICENTEPRL QDKQKGAQTFEEMNTIIPDLIVKIGETKIESIRLLFDGLQPVNLKQLTYYLLDI VIQELPPELAKVKQKETSMTSM"
REFERENCE	4	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kueh, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carrincci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayaashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851	polya_signal	3044. 3049 /note="putative"
REFERENCE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 3064)	polya_site	3064 /note="putative"
REFERENCE	6	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayaashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoch, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayaashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayaashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome	BASE COUNT	886 a 637 c 704 g 837 t
REFERENCE	7	Alignment Scores: Pred. No.: 7,596-245 Length: 3064 Score: 2347.00 Matches: 446 Percent Similarity: 99.56% Conservative: 7 Best Local Similarity: 98.02% Mismatches: 2 Query Match: 96.31% Indels: 0 DB: 11 Gaps: 0	ORIGIN	
REFERENCE	8	US-09-744-313A-1 (1-465) x AK049371 (1-3064)		
REFERENCE	9	QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIleIleLys 30 Db 1570 AGAACACACAAAAGAGGAGATCGTTTGGATCAGCAGATAGTAGCAATAATATAA 1629		
REFERENCE	10	QY 31 GlyValPheLysSerThrThrMetGluClyAlaMetLeuProAsnTyrGlyValalaGlu 50 Db 1630 GCGCGTATTCAAGAGTACGACCAATGAGGAGCTGTGCTGCTTAATTCAGGGGTGCTGAA 1689		



Qy 51 GlyGluAspPheIleGluGluGlyLeuValMetGluAspSerProValGlu 70  
Db 1690 GCGCAAGATGACTTTATTGAAGAGGATTTGTGTAATGAGGATGACTCTCCAGTAGAA 1749  
Qy 71 AlaValSerThrProAsnThrProArgSerLeuAlaTLPLeuIleSerIleProTyr 90  
Db 1750 GCTGTGAGCACACCTTAACACTCTCGAAACCTTGTCTGCTGGAATAATAGCATTTCCGAT 1809  
Qy 91 ValAspPheGluAspProSerSerGluArgGlyGluValLysLysGluArgIleProVal 110  
Db 1810 GTAGACTTTTGAAGATCCCTCTCTGAAAGAAAGAGAGAGAGAGATTCCTGTG 1869  
Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyr 130  
Db 1870 TTTTGTATCATGTTGAAAGAAACACACAGCAGCAGTGTGACAGCGCTGAGCATTTG 1929  
Qy 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
Db 1930 TCTGCTATAGAGATATCTTGAATCTATGTTCTTGAATCAAACTAACAGATTTTCAT 1989  
Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
Db 1990 GGCACATTTCTGATGCTCAGCTTCATCCAAAGATCATTTGGCCCAAAACTATGAG 2049  
Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 2050 TTTTGAAGTCGAGAGAGAGAGAGTTCCAGAGTATCTCGAAACTTTGTGCGACCCG 2109  
Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
Db 2110 GAGCTGAGTAACAGCAGCTTCTGCTGACTTTCTCTCCCCAACCGTGGGAGACAG 2169  
Qy 211 PheLeuAspLysIleLeuProAspValLeuLeuGlyLysIleIleLysSerValProGly 230  
Db 2170 TTTCTAGATAAGATATCTCCAGATGAATCTTGGAAAAATTAATAATCTGTTCTCTGG 2239  
Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
Db 2230 AACTAATGAAGAGAGAGAGTTCAGATTTGGAGCATTCATCATGAGTTTCAATATCT 2289  
Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 2290 TCGCAATCTCCCAACCTTAACCGAGTGGCCAGAACTGACGATTTCTCAGCCCACTTCC 2349  
Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
Db 2350 GAAACCAACAGAAGCTTTTCAACGATCTGTTTAAGAATAATGCAAAACCGCGCTGAGA 2409  
Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 2410 ACAGAGCGAAAGCAAAATCAGAACTACTTTCATGGAGGTGATGCTGTGGAGGAGTCTAT 2469  
Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHisLeu 330  
Db 2470 GATTACCTGATGATGAGGCGAGTAGTTTTTCAAGTCCAGAGTGGCTTCATCATCTT 2529  
Qy 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
Db 2530 CTAATGGGAACCTCGAATCTCTTTAAGAACACCTTGAAGATGTACACAGACTACTACCTC 2589  
Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
Db 2590 CAGTGCAGCTGAGCAGCTGTTTTCAGGAGCAGCGGCTGGTCTCCCTCATCACCTTCTC 2649  
Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
Db 2650 AGAGATGCTATATTTTGGAAAAATATGAACTCGCTCTCTCCAGATGAGCAAAAGGA 2709  
Qy 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
Db 2710 GCAAAACAGACTTTTGAAGAAATGATGAATTACATTACATCTCCAGATCTGATGATCAAGTATC 2769

411 GlyGluThrThrIleTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
2770 GGTGAAGAACCAGTACCAAGACATCACTCTCTGTTGATGCTGCGAGCAGCCAGTG 2829  
431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuLeuPheProGlu 450  
2830 CTCAACAGCAGCTGACCTATGTTGCTGCGATGTTGATACAGAGCTTTTCCAGAG 2889  
451 LeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
2890 CTAATAAGGTACAAAGAGAGAGCTACCTCCATGACATCTCTGGATG 2934  
RESULT 2  
AK082389 2634 bp mRNA linear HTC 05-DEC-2002  
LOCUS AK082389  
DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:C230045H05 product:HYPOPHYSICAL 102.9 KDA  
PROTEIN (SORTING NEXIN 14) homolog [Homo sapiens], full insert  
sequence.  
ACCESSION AK082389  
VERSION AK082389.1 GI:26100628  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gotobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehli, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,  
Quackenbush, J., Schriml, L.M., Stauble, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bona, M., Brownstein, M.J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, H., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohetsuki, S.  
and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection

Nature 409 (6821), 685-690 (2001)

21085660

11217851

5

THE FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2634)

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hangaki,T., Hara,A., Haseizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp, URL:htp://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL:htp://genome.gsc.riken.go.jp/

URL:htp://fantom.gsc.riken.go.jp/.

FEATURES

source

1..2634

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="PANTOM.DB:C230045H05"

/db\_xref="taxon:10090"

/clone="C230045H05"

/tissue\_type="cerebellum"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stages="0 day neonate"

misc\_feature

1..2634

/note="HYPOTHETICAL 102.9 KDA PROTEIN (SORTING NEXIN 14) homolog (Homo sapiens) [SPTR|Q9BSD1, evidence: FASTV, 93.3%id, 99.6%length, match=2781]"

BASE COUNT

720 a 570 c 524 g 720 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	2634
Score:	1949.00	Matches: 369
Percent Similarity:	99.73%	Conservative: 5
Best Local Similarity:	98.40%	Mismatches: 1
Query Match:	79.98%	Indels: 0
DB:	11	Gaps: 0

US-09-744-313A-1 (1-465) x AK082389 (1-2634)

Qy 10 PheArgAsnThrGlnLysArgGlyClnSerPheGlyIleSerArgIleGlySerIlySile 29

Db 1509 TTCGGAGGACACACAAAAGAGGAGCATCGTTTCGATCAGCAGATAGGTAGCAAAATT 1566

Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrglyValAla 49

**DEFINITION** EX372281 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
Homo sapiens cDNA clone CSODL007YE10 3-PRIME, mRNA sequence.  
**ACCESSION** EX372281  
**VERSION** 1  
**KEYWORDS** EST.

**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 917)

**AUTHORS** Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished  
**COMMENT** Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 3600.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/ccluster.cgi?seq=CSOBAI04ZB11.CS04210\_1&cluster=3600.r.  
Contact : Feng Liang Email : f.liang@lifetech.com URL :  
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CSOBAIO4ZB11\_CS04210\_1.  
Location/Qualifiers  
1..917  
    /organism="Homo sapiens"  
    /mol\_type="mRNA"  
    /db\_xref="taxon:9606"  
    /clone="CSODL007YE10"  
    /cell\_type="B CELLS (RAMOS CELL LINE)" COT 25-NORMALIZED"  
    /cell\_line="RAMOS CELL LINE"  
    /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
       25-NORMALIZED"  
note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPOT 6 vector. Library was normalized."  
BASE COUNT   257 A   168 C   177 G   310 T   5 others  
ORIGIN

**Alignment Scores:**

Pred. No. :	6.6e+155	Length:	917
Score:	1517.00	Matches:	293
Percent Similarity:	96.72%	Conservative:	2
Best Local Similarity:	96.07%	Mismatches:	9
Query Match:	62.25%	Indels:	1
DB:	13	Gaps:	0

US-09-744-313A-1 (1-465) x BX372281 (1-917)

```
Qy      135 ArgTyrrLeuGluPheTrValLeuGluserylserlyseuthrcluphehisglyalaphePro 15
       ::::::::::::::::::::|:::|||||||:::~::~:
Db      915 AAGATTCTTGAAATCCTTAGACTGAATAACAATACTAACAGAATTTCATGGCGCATTTCCCT 85
                               |||||
Qy     155 AspaladinLeuproSerLySArglleilegylprolysasntrycglupeulelysser 17
                                   |
Db      855 GATGCCCAAGCTCCCCTCTTAAGAGATCATTTGGCCCCAAAAAATATGTGNATCTTAAGTCNA 79
                                     ~~~~~
Qy     175 LysargGluGluPheGlnGinufyrLeuGlnLysLeuGlnHisProGluLeuSerAsn 19
                                   |
Db      795 GNNAAGGAAGAGTTCCCAAGATATCTACAGAACTCTGCAGCATCCAGAACTGAGTAAT 73
                                         |
Qy     195 SerGlnLeuAlaAsppHeLeuSerProasnglyclythrGlnpheleuasplys 21
                                   |
Db      735 AGTCAAACCTCTGGCAGACTTCTTTCTCCCTTAATGGTGGGAAAACACAAATTTCTTGATAAG 67
                                           ||
Qy     215 IleleuProaspValasnLeuglylvyleilleleyservvalprodglylsleumetylvs 23
                                   |
Db      675 ATACTACAGAGTPAATCTCTGGGAAATATTANAATCTGTTCTCTGGAAACATAATGAAA 61
                                               |
Qy     235 GlulysglyGlnHisLeuGluPuropheileMetAsnPheileasnSerCySGlusSerPro 25
```

```

/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30243103"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 164"
/notes="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Non-normalized full-length enriched library from
pooled mouse embryonic limb, maxilla and mandible, day
10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp, Average insert size 1.8k bp. Priming
sequence: 5'GACTGTTCTAGATCGGAGCGCCGCC(T) 3'. Tissue
contributed by, David Rowe. Library constructed by ResGen,
Invitrogen Corp."
Invitrogen Corp."

BASE COUNT 276 a 207 c 194 g 226 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 1-27e-142 Length: 905
Score: 1404.50 Matches: 277
Percent Similarity: 95.61% Conservatives: 6
Best Local Similarity: 93.58% Mismatches: 9
Query Match: 57.63% Indels: 4
DB: 14 Gaps: 1

US-09-744-313A-1 (1-465) x CB182722 (1-905)

Qy 143 GluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArg 162
Db 12 GAATCAAACTAAAGAAATTCATGTCACATTTCTCTGATGCTCAGCTCCATCCAAAGG 71
Qy 163 IleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyr 182
Db 72 ATCATTTGGCCCCAAAACATGATGTTTTGAATCGAAGAGAGAGAGATTCAGGAGTAT 131
Qy 183 LeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaPheLeu 202
Db 132 CTGCAGAAATTTGTGCGAGCACCCGAGCTGAGTAACAGCAGCTTCTGCTGACTTC 191
Qy 203 SerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGly 222
Db 192 TCCCCCAACGGTGGGGAGACACAGTTCTTAGATAAGATCTCCAGATGTAAATCTTGA 251
Qy 223 LysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluPro 242
Db 252 AAAATTTAAAGTCTGTCTCTCGGAAACTAATGAAAGAGAAAGGTCCAGATTTCGAGC 311
Qy 243 PheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGlu 262
Db 312 TTCATCATGAGTTTCATTAATCTTGGCAATCTCCCAACCTAAACCGAGTCGGCCAGAA 371
Qy 263 LeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLys 282
Db 372 CTGACGATTTCTAGCCCCCATCTCCGAAAACACACAGAGGTTTTCAACGATCTGTTTAA 431
Qy 283 AsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGlu 302
Db 432 AATATGCAACCGCGCTGAGAACACACAGAGCGAAAGCAAAATCAGAACTACTTCATGGAG 491
Qy 303 ValMetThrValGluGlyValTyrAspTyrIleuMetTyrValGlyArgValValPheGln 322
Db 492 GTGATGACTGTGGACGGAGTCTATGATTAATCTGATGTATGAGCGAGTAGTTTTCACAA 551
Qy 323 ValProAspTyrPheLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu 342
Db 552 GTCCAGACTGGCTTCATCTTCTTAATGGAACTCGAATCTCTTTAGAACACACCTTG 611
Qy 343 GluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg 362
Db 612 GAAATGTACACAGACTATTACTTCAGTGAAGCTGGAGCAGCTGTTTTCAGGAGCACCG 671
Qy 363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArg 382

```

```

672 CTGGTCTCTCCATCATCCCTTCTCAGAGATGCTATATTTTGTGGAATACTGAACTCGC 731
383 SerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIle 402
732 TCTCTCNAGATAGCAAAAGAGAGCAAAACAGACTTTTGGAAATGATGATTACATT 791
403 ProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeu 421
792 CCAGATCTGATAGTCAAGTGTATCGGTGAAGAAACCCAGTACGAAAGCTTCAGAACT 851
422 LeuPheAspGlyLeu---GlnGlnProValLeuAsnLysGln 434
852 CTGTTGATGGGCTCGCAGCAGCCAGTTCCTCAACAGCAG 893

RESULT 5
AL574298/c 1043 bp mRNA linear EST 01-JUN-2003
DEFINITION AL574298 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI057YD12 3-PRIME, mRNA sequence.
ACCESSION AL574298 GI:31312617
VERSION AL574298.2
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1043)
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 16, 2001 this sequence version replaced gi:12934370.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3600.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODI057YD12&cluster=3600.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI057YD12B06NP1.

FEATURES
Location/Qualifiers
1..1043
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI057YD12"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(GT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 307 a 196 c 182 g 338 t 20 others
ORIGIN

Alignment Scores:
Pred. No.: 2-04e-136 Length: 1043
Score: 1348.50 Matches: 273
Percent Similarity: 91.39% Conservatives: 3
Best Local Similarity: 90.40% Mismatches: 23
Query Match: 55.33% Indels: 5
DB: 9 Gaps: 1

US-09-744-313A-1 (1-465) x AL574298 (1-1043)

Qy 165 GlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluLeuGln 184
Db 1041 GGCCCCAAATATGATGATTTT-----AAAKYCAARGGGAGGKGTCCAAGATWTCAWC 988
Qy 185 LysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerPro 204

```



```
QY 245 MetAsnPhelIeAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr 264
Db 363 ATGATTTATATCAATTCCTCGGAATCTCCAGACCTAAACCCAGCAGCCCTGAGCTTACG 422
QY 265 IleLeuSerProThrSerGluAsnAsnLysLysLysLysLysLysLysLysLysLysLys 284
Db 423 ATTCTGAGTCCCACTCTCTGAGAACACAGAGAGCTTTTCAATGATTTATTCAGAAATAAT 482
QY 285 AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMet 304
Db 483 GCAACCCGCTCTGAGAAATACAGAAAGACGACAAATACAGAACTCTTCAATGATTTATTCAGAAATAAT 542
QY 305 ThrValGluGluValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPro 324
Db 543 ACTGTAGAGGAGTCTATGACTACTTATTTGATGTGTGTAGAGTGTCTTCCACATCTCT 602
QY 325 AspTrpLeuHisLysLeuMetGlyThrArgLysLysLysLysLysLysLysLysLysLys 344
Db 603 GACTGCTTCATCATCTCTGATGGAGGAGAAATCTCTTCAAAATACATTCGAACTG 662
QY 345 TyrThrAspTyrTyrLeuGlnCysLysLysLysLysLysLysLysLysLysLysLys 364
Db 663 TACACGACTATTTACTTGCATTTATAGTTAGAACAGCTATGTGAGGAGCAGCCGTTTGGTT 722
QY 365 SerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeu 384
Db 723 TCTCTCATCAACACTGCTGAGAGATGCTGTGTCTGTGAGATCTGAAACCCGCTCTGTC 781
QY 385 GlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMet---MetAsnTyr-IlePr 403
Db 782 CAGGACAACAGAGAGAGCAAGCAAGACTTTTTCGAAAGAACTGATGAGATACCTCTCC 841
QY 403 AspLeuLeuValLysCysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 422
Db 842 AGATTTTATAGGCAAGTGCATTTGGGGGAGAGAGCTAAATACGAAAGCTCTCCAGGCTTC 901
QY 422 euPhe-----AspGlyLeuGlnProValLeuAsnLysGlnLeuThr---TyrValL 439
Db 902 TGTTTGTAGCTGCTGCAGCAGCCACCCAGTGGCTTCAACGACAGTTGACTTAACTGCT 961
QY 439 euLeuAspIleValIle-GlnGluLeuPhePro 449
Db 962 GTTGGGACATCGGATTCAGGAATCCTTCCA 994

RESULT 7
BG191447 818 bp mRNA linear EST 21-APR-2001
RST10541 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.
BG191447
BG191447.1 GI:13713134
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 818)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 428.
```

```
FEATURES
source
1..818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
```

```
BASE COUNT 275 a 159 c 150 g 229 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 4,23e-128 Length: 818
Score: 1271.00 Matches: 259
Percent Similarity: 95.97% Conservative: 3
Best Local Similarity: 94.87% Mismatches: 11
Query Match: 52.15% Indels: 5
DB: 10 Gaps: 0
```

```
US-09-744-313A-1 (1-465) x BG191447 (1-818)
```

```
QY 124 GlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGlu 143
Db 3 GGACACAGAGCTGA-CATTGGTCTGCTATAGAGATATCTGAAATCTATGATTTGAA 61
QY 144 SerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArgile 163
Db 62 TCAAACTAACAGAAATTCATGTCGTCATTTCTGATGCCAGCTTCTCTTAAGAGATC 121
QY 164 IleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluThrLeu 183
Db 122 ATTGGCCCCAAAATTTATGAATTTCTTAAAGTCAAAGAGGAGAGTTCCAAGATATCTA 181
QY 184 GlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSer 203
Db 182 CAGAAACTTCTGCAGCATCCAGAACTGAGTAATAGTCAACTCTCGCAGACTTCTTTTCC 241
QY 204 ProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLys 223
Db 242 CCTAATGTGGGGAACACACAAATTTCTTGATAGATATCTACCAGATGAATCTTGGAAA 301
QY 224 IleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPhe 243
Db 302 ATTATAAATCTGCTCTCTCGAAAACCTAATGAAAGAGAAAGGTCAAGCATTTGGAACCTTT 361
QY 244 IleMetAsnPhelIeAsnSerCysGluSerProLysProLysProSerArgProGluLeu 263
Db 362 ATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTG 421
QY 264 ThrIleLeuSerProThrSerGluAsnAsnLysLysLysLysLysLysLysLysLysLys 283
Db 422 ACCATTTCTGACCCCTACTCTCAGAAAACACACAGAGACTTTTCAATGATCTGTTTAAAT 481
QY 284 AsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluVal 303
Db 482 AATGCANACCGTCTGTAANATAACAGAGAGAAACAAATATTTATTTATGAGAGTGTG 541
QY 304 MetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnVal 323
Db 542 ATGACTGTAGAGGAGTCTATGATTTACCTGATGTATGTAGGACGGGTAGTTTTCAGGTT 601
QY 324 ProAspTrpLeuHisLysLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGlu 343
Db 602 CCTGACTGGCT-CATCATCTCTTAAATGGGAACTCGAATCTCTTNTAAAAACACCCCTGGAA 660
QY 344 MetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlnHisArgLeu 363
Db 661 ATGTATACTGATTTACTATCTTCTGATGTAACTAGACAGCTATTTTCAGGAGCACCTTT- 719
```

Qy 364 ValSerLeuThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSer 383  
 Db 720 GTCTCACTATCAACTTCTCAAGAGTGTATATTCTGTGAACACACCTGGCTCT 779  
 Qy 384 LeuGluAspGlnGlyGlyAlaLysGlnThrPheGlu 396  
 Db 780 CTCAGAAATAG-CAAAAGGAGC-AAACAGACTTTNGAG 816

RESULT 8  
 BQ053127  
 LOCUS  
 DEFINITION  
 AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711  
 5', mRNA sequence.  
 ACCESSION  
 BQ053127  
 VERSION  
 BQ053127.1 GI:19812467  
 KEYWORDS  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1096)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCW2121 row: e column: 16  
 High quality sequence stop: 675.

FEATURES  
 source  
 1..1096  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5934711"  
 /tissue\_type="natural killer cells, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_106"  
 /notes="Organ: blood; Vector: pOTB7; Site: 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."  
 BASE COUNT 355 a 193 c 226 g 305 t 17 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,26e-127 Length: 1096  
 Score: 1266.00 Matches: 243  
 Percent Similarity: 99.18% Conservative: 0  
 Best Local Similarity: 99.18% Mismatches: 2  
 Query Match: 51.95% Indels: 0  
 DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x BQ053127 (1-1096)

Qy 221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGlyGlnHisLeu 240  
 Db 7 TTGGGAAATTTAAATCTGTCTCGAAACTAATGAAGAGAGGTCAGCATTTG 66  
 Qy 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
 Db 67 GAACCTTTTATCATGAATTTTCATTAATCTGTGAGTCTCCAAAGCCTAAACCAAGTAGA 126

Qy 261 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280  
 Db 127 CCAGAACTGACCACTTCTCAGCCCTACTTTCAGAAAACAACAAGAGCTTTTCAATGATCTG 186  
 Qy 281 PheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnThrPhe 300  
 Db 187 TTTAAATAATATGCAACCGTGTGTAATATACAGAGAGAAAGCAAAATCAGAAATTTATTT 246  
 Qy 301 MetGluValMetThrValGluGlyValTyAspTyLeuMetTyValGlyArgValVal 320  
 Db 247 ATGGAGGTGATGACTGTAGAAGAGGTCTATGATTACTGATGATGTAGGACGGGTAGTT 306  
 Qy 321 PheGlnValProAspTrpLeuHisLeuMetGlyThrArgIleLeuPheLysAsn 340  
 Db 307 TTCCAGGTTCTGACTGGCTTCATCTCTTAATGGAACTCGAATCTCTTTTAAAAAC 366  
 Qy 341 ThrLeuGluMetTyThrAspTyTyLeuGlnCysLysLeuGlnLeuPheGlnGlu 360  
 Db 367 ACCCTGAAATGTATATCTGATTAATCTTCACTGTAAACTAGAACAGCTATTTCAGGAG 426  
 Qy 361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380  
 Db 427 CACCGTTTGGTCTCACTCATACACATCTCTCAGAGATGCTATATCTGTGAAAACACATGAA 486  
 Qy 381 ProArgSerLeuGlnAspLysGlnLysGlyValAlaLysGlnThrPheGluGluMetMetAsn 400  
 Db 487 CCTCGCTCTCTCCAGATAGCAAAAGAGGCAAAACAGACTTTTGAAGAAATGATGAAT 546  
 Qy 401 TyrlleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyGluSerIleArg 420  
 Db 547 TACATTCAGATCTGTGTAGTCAAGTGATTGGTGAAGAAACCAAGTATGAAGACATCAGA 606  
 Qy 421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyValLeuLeu 440  
 Db 607 CTTCCTGTTGATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTG 666  
 Qy 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
 Db 667 GACATTGTGATACAGGAACCTTTCCAGACTCAATAGGTACAAAGGGAAGTACTCTCT 726  
 Qy 461 ValThrSerTrpMet 465  
 Db 727 GTGACATCTTGGATG 741

RESULT 9  
 BQ230190  
 LOCUS  
 DEFINITION  
 AGENCOURT\_7567151 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:6054428  
 5', mRNA sequence.  
 ACCESSION  
 BQ230190  
 VERSION  
 BQ230190.1 GI:20411590  
 KEYWORDS  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 921)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCW3313 row: a column: 21  
 High quality sequence stop: 639.  
 Location/Qualifiers

FEATURES



```
source
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6054428"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 70"
/notes="Organ: pancreas; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."
BASE COUNT      307 a 168 c 181 g 265 t
ORIGIN
Alignment Scores:
Pred. No.:      1,33e-123      Length:      921
Score:          1230.50      Matches:    240
Percent Similarity: 96.41%      Conservative: 2
Best Local Similarity: 95.62%      Mismatches: 2
Query Match:    50.49%      Indels:    7
DB:             13          Gaps:      2

US-09-744-313A-1 (1-465) x BQ230190 (1-921)
QY      221 LeuGlyLysIleLeuValProGlyLysLeuMetLysGluLysGlnHisLeu 240
DB      5 TTGGGGAATTAATAAAATCTGTTCTGGAAATTAATGAAGAGGTCAGCATTTG 64
QY      241 GluProPheIleMetAsnPhelIleAsnSerCysGluSerProLysProLysProSerArg 260
DB      65 GAACCTTTTATCATGAATTTCAATTAATCTTGAGTCTCCAAAGCCCTAAACCAAGTAGA 124
QY      261 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280
DB      125 CCAGAACTGACCATTTCTAGCCCTACTTCCAGAAAACACAGAACGCTTTTCAATGATCTG 184
QY      281 PheLysAsnAlaAsnArgAlaGluAsnThrGluArgCysGlnAsnGlnAsnTyrPhe 300
DB      185 TTTAAAAAATGCAACCGTGTGAAATATACAGAGAGAAACAAATCAGANTATTATTT 244
QY      301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320
DB      245 ATGGAGGTGATGCTGTAGAGGATCTATGATTACTTATGATGTGTAGGACGGGTAGT 304
QY      321 PheGlnValProAspTrpLeuHisLysLeuMetGlyThrArgIleLeuPheLysAsn 340
DB      305 TTCCAGGTTTCCTGACTGGCTTCATCATCTTAATGGGAACCTCGAATCCTCTTTAAAAAC 364
QY      341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu 360
DB      365 ACCCTGGAATGTATCTACTGATTAATCTCTAGTGTAAACTAGAACAGCTATTTCCAGGAG 424
QY      361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380
DB      425 CACGTTTGTCTCATCTCATACCTTCTCAGAGATGCTATATTCTGTGAAAACACTGAA 484
QY      381 ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400
DB      485 CCTCGCTCTCTCCAGATAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAAT 544
QY      401 TyrIleProAspLeuLeuValLysCysIleGlyGluThrLysTyrCysIleSerIleArg 420
DB      545 TACATTCAGATCTGTTAGTCAAGTGTATTGGTGAAGAAACCAAGATGATGAAGCATCAGA 604
QY      421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440
DB      605 CTTCGTGTTGATGGCTTACAGCAACCAAGTACTCAACAGAGCTGACTTATGTTTATTG 664
QY      441 AspIleValIleGlnGluLeuPheProCysGluLeuAsnLysValGlnLysGlu----- 457
DB      665 GACATTGTGATACAGGAACTGTTCCAGAGCTCCATAGGTACAAAAGGAAGTTAACT 724
```

```
QY      458 ---ValThrSerVal-----ThrSerTrp 464
DB      725 CTGGTGACATCTTTGGGATGTAAACACACCTGG 757

RESULT 10
AL578668/c      903 bp      mRNA      linear      EST 01-JUN-2003
LOCUS      AL578668 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION      cDNA clone CS0DK004YA04 3-PRIME, mRNA sequence.
ACCESSION      AL578668
VERSION      AL578668.1 GI:12942962
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 903)
AUTHORS      Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 131 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3600.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK004BA02NP1cluster=3600.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DK004BA02NP1.

FEATURES             source
1..903
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK004YA04"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      272 a 160 c 159 g 290 t 22 others
ORIGIN
Alignment Scores:
Pred. No.:      1.82e-122      Length:      903
Score:          1220.00      Matches:    239
Percent Similarity: 94.49%      Conservative: 1
Best Local Similarity: 94.09%      Mismatches: 14
Query Match:    50.06%      Indels:    1
DB:             9          Gaps:      0

US-09-744-313A-1 (1-465) x AL578668 (1-903)
QY      212 LeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValProGlyLys 231
DB      902 CTTGATTAAGTACTTACCAGATGTAAATCTTGGAAATATATAAAATCTGTTCTCGGAAA 843
QY      232 LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPhelIleAsnSerCys 251
DB      842 CTAATGAAGAGAGAGGTCAGCATTTTGAACCTTTTATCATGAATTTCAATTAATCTTGT 793
QY      252 GluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGlu 271
DB      782 GAGTCTCCAAAGCCCTAAACCAAGTAGCCAGACTGACCTTTCTAGCCCTACTTACGAA 723
QY      272 AsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThr 291
DB      722 AACACAGAGAGGCTTTTCATGATCTGTTTAAAAAATAATGCAACCGTGTGAAATACA 663
```





QY 362 rgluValSerleuThrleuLeuArgAspAlaile-PheCysGluuAnThrGluPro 381

Db 620 CGTAGGTTCTCTCAACACCTGCTGAGAGATGCTGTGGTTCTGTGGAATACGACCG 679

QY 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401

Db 680 CGCTCTGTCCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 739

QY 402 IleProAspLeuValLysCysIleGlyGluGlu-----ThrLysTyrGluSer 418

Db 740 ATCCAGATTATAGGCAAGTGCCTGGGGAAGAGAGGCTAACTACGAAG-----GGC 793

QY 419 IleArgLeuLeuPheAsp---GlyLeuGlnGlnPro---ValLeuAsnLys-GlnLeuTh 436

Db 794 ATCAGGCTTCTAGTTCGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGC 853

QY 436 rTyrValLeuLeuAspIle-ValIleGlnGlu-LeuPheProGlu 450

Db 854 TTACGCTGCTGTGGACATCGGGANTTCAAGAAACTCTTTCCAGAA 898

RESULT 12

LOCUS BG256631 1105 bp mRNA linear EST 13-FEB-2001

DEFINITION 602370809F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4478858 5', mRNA sequence.

ACCESSION BG256631

VERSION BG256631.1 GI:12766447

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NTH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM10310 row: a column: 03  
High quality sequence stop: 645.  
Location/Qualifiers  
1. .1105  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4478858"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_hosts="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC Library."

BASE COUNT 380 a 235 c 208 g 282 t

ORIGIN

US-09-744-313A-1 (1-465) x BG256631 (1-1105)

QY 217 ProAspValAsnLeuGlyLysIleLeuValProGlyLysLeuMetLysGluLys 236

Db 1 CCAGATGTAAATCTTGGGAAATTTAAATCTGTTCTCGGAAACTAATGAAAGAGAA 60

QY 237 GlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysPro 256

Db 61 GGTTCAGCATTTGGAACTTTTATCATGAATTTCAATTAATCTTGTGAGTCTCCAAAGCCT 120

QY 257 LysProSerArgProGluLeuThrIleLeuSerProThrSerGluuAnLysLysLeu 276

Db 121 AAACCAAGTAGACAGACAGCACTTCTCAGCCCTACTTCAGAAACCAACAGAAAGCTT 180

QY 277 PheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsn 296

Db 181 TTCAATGATCTGTTTAAAAATAATCAAAACCCGCTGCTGAAATAACAGAGAAAGCAAAAT 240

QY 297 GlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal 316

Db 241 CAGAAATATTTTATGGAGGTGATGCTAGAGAGAGTCTATGATTTACTCTGATATGTA 300

QY 317 GlyArgValValPheGlnValProAspTyrPheHisLeuLeuMetGlyThrArgIle 336

Db 301 GGACGGGTAGTTTCCAGGTTCTGACTGCTTTCATCATCTCTTAATGGAACTCGAATC 360

QY 337 LeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGln 356

Db 361 CTCTTTAAAAACACCCCTCGAAATGTATCTACTTACTTATCTTCTCAGTGTAAACTAGAACAG 420

QY 357 LeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCys 376

Db 421 CTATTTTCAGAGACACCGTTTGGTCTCACTCATTAACCTTCTCAGAGATGCTATATTTCTGT 480

QY 377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396

Db 481 GAAACACTGACCTGCTCTCTCCAGATAAGCAAAAGAGCAAAACAGACTTTTGA 540

QY 397 GluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr 416

Db 541 GAAATGATGAATTACATTCAGATCTGTAGTCAAGTGTATTGTTGTAAGAAACCAAGTAT 600

QY 417 GluSerIleArgLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436

Db 601 GAAAGCATCAGACTTCTGTT-GATGGCTTACAGCAACCACTACTCAACAGCAGCTGACT 659

QY 437 TyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysValGln-L 456

Db 560 TATGTTTTATCGGACTTGGTGTATCCGGGAACCGTTCCAGAGATTCAATAGGTGACACA 719

QY 456 ysGluValThrSerValThrSerTyr 464

Db 720 CGGAAGTTACTCTCTGTGGACATCTGG 745

RESULT 13

LOCUS BG293966 1093 bp mRNA linear EST 21-FEB-2001

DEFINITION 602391002F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:4502737 5', mRNA sequence.

ACCESSION BG293966

VERSION BG293966.1 GI:13054129

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 1093)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 cDNA Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10372 row: d column: 02  
 High quality sequence stop: 738.  
 Location/Qualifiers

# FEATURES

```

source
1..1093
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4502737"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 94"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH MGC Library."
BASE COUNT 353 a 223 c 280 g 237 t
ORIGIN
Alignment Scores:
Pred. No.: 4,19e-117 Length: 1093
Score: 1172.00 Matches: 246
Percent Similarity: 86.15% Conservative: 9
Best Local Similarity: 83.11% Mismatches: 34
Query Match: 48.09% Indels: 8
DB: 10 Gaps: 3
US-09-744-313A-1 (1-465) x BG293966 (1-1093)
QY 19 SerPheGlyLeuSerArgIleGlySerLysIleLysGlyValPheLysSerThrMet 38
Db 3 TCCTTTGGATCAGCAGATAGGTAGCAAAATTAAGCGCTATTCAAGGTACGCAATG 62
QY 39 GluGlyAlaMetLeuProAsnTyrglyValAlaGluGlyGluAspPheIleGlu 58
Db 63 GAGGGAGCTGTGCTCTTAATACGGGTGGCTGAAGGGAAGATGACTTTATTAGAA 122
QY 59 GlyIleValValMetGluAspSerProValGluAlaValSerThrProAsnThrPro 78
Db 123 GCGATTGTGTAATGAGGATGACTCTCCAGTAGAGCTGTGAGCACACCTAACACTCT 182
QY 79 ArgAsnLeuAlaAlaTrpLysIleSerIleProTyrglyValAspPheGluAspProSer 98
Db 183 CGAAACCTTGCTCGCTGGAAAATTAGCATTCGTATGTAGACTTTTGAAGATCCCTCC 242
QY 99 SerGluArgLysGlyLysLysGluArgIleProValPheCysIleAspValGluArgAsn 118
Db 243 TCTGAAGGAAAGAGAGAGAGAGATTCCTGTTTGTATCGATTTGAAGAAAC 302
QY 119 AspArgAlaValGlyHisGluProGluHisIleTrpSerValTyrgargTyrglyLeu 138
Db 303 GACAGACGACGAGCTTGACACGAGCTGAGCACTTGCTGTCTATAGAGATCTTGA 362
QY 139 PheTyrglyLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeu 158
Db 363 TTCTATGTTCTTGAATCAAACTTAACAGAAATTTATGGCACATTTCTGATGCTCAGCT 422
QY 159 ProSerLysArgIleIleGlyProLysAsnTyrglyPheLeuLysSerLysArgGlu 178
Db 423 CCATCCAAAAGGATCATTTGGCCCCCAAAACTATGAGTTTGAAGTCGAAGAGAGAG 482
QY 179 PheGlnGlyTyrglyLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeu 198
Db 483 TTCAGAGATTTCTGAGAAACTTTGTCAGACCCCGAGCTGAGTAAACGACCTTCG 542
QY 199 AlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAsp 218

```

```

Db 543 GCTGACTTCTCTCCCAACGGTGGGAGACACAGTTTCTAGATAAGATATCTCCAGAT 502
QY 219 ValAsnLeuGlyLys-IleIle-LysSerValProGlyLysLeu-MetLysGlyLys-Gl 237
Db 603 GTAAATCTTGGAAACATTATAACAGTCTGTCTCTGGGAACTAATTGAAGAGACGG 562
QY 237 YGlnHisLeuGluProPheIleMetAsnPhellehenSerCysGluSerProLysProly 257
Db 663 TCAGCATTCGGAGCCATTTCATCATGAGTTTCTTATTTCTTGCATCTCCAAAATAA 722
QY 257 sProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPh 277
Db 723 CGAGTCTG---GCAGAACTGGGGAAATCTCAGGCCAATTCGAA---ACACAAGAGCTTT 776
QY 277 eAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGl 297
Db 777 CCAGACCGGCTTAGCAATAA---GGGACCGGGCGGAGAAACCGAGGGGGAACCAATGA 835
QY 297 nAsnTyrglyPheMetGluValMetThrValGluGlyValTyrgly 310
Db 836 A---ATACTCAGGAGAGGAAACGGGGAGCGGGCGCTAC 872
RESULT 14
BU273460 754 bp mRNA linear EST 26-NOV-2002
LOCUS 603375273F2 CSFQCHN53 Gallus gallus cDNA clone ChEST288a20 5', mRNA
DEFINITION
sequence.
ACCESSION BU273460
VERSION BU273460.1 GI:25544410
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 754)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
12445392
PUBMED
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..754
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="ChEST288a20"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSFQCHN53"
/note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS

```

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

BASE COUNT 240 a 149 c 168 g 197 t  
ORIGIN

## Alignment Scores:

Pred. No.: 3,06e-116 Length: 754  
Score: 1162.00 Matches: 228  
Percent Similarity: 90.49% Conservative: 10  
Best Local Similarity: 86.69% Mismatches: 12  
Query Match: 47.68% Indels: 14  
DB: 13 Gaps: 1

US-09-744-313A-1 (1-465) x BU273460 (1-754)

Qy 163 lilelleglyprolysaenlyrghupheleuylserylsaragglupheglngluYr 182  
Db 4 ATCATTTGGCCCAAGAACTATGATTCCTTAATCCAAAGGAGGAGTTCCAGAGTAT 63  
Qy 183 LeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeu 202  
Db 64 CTACAGAACTTCTGCAACTCCAGAGCTAAGTAACAGCCACTTTTAGCTGACTTCTG 123  
Qy 203 SerProAsnGlyGluThrGlnPheLeuAspLyslleLeuProAspValAsnLeuGly 222  
Db 124 TCCCTTAATGGAGGAGACTCAGTTCTTGAT----- 156  
Qy 223 Lysllellelye-SerValProGlyLysLeuMetLysGluLysGlnHisLeuGluPr 242  
Db 157 -----AAATCTGTTCAGGAAGCTGTGAAGAGAAAGTCACTTTGGGCC 207  
Qy 242 oPhelleMetAsnPhaileAsnSerCysGluSerProLysProLysProSerArgProG1 262  
Db 208 ATTCAATCATGAATTTATCAATTCCTGCGAATCTCCAGAGCTTAACCGAGAGCCCTGA 267  
Qy 262 uLeuThrilleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLy 282  
Db 268 GCATTACGATTCTCAGTCCCACTTCTGAGAACACAGAGAGCTTTCAATGATTATTCAA 327  
Qy 282 sAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetG1 302  
Db 328 GAATAATGCAACCGGCTCTGAGAAATACAGAAAGACGCAAAATCAGAACTTTCATGGA 387  
Qy 302 uValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyValValPheG1 322  
Db 388 AATGATGACTGTAGAGAGGCTCTATGACTACTTATTGTATGTGGTAGAGTGTCTTCCA 447  
Qy 322 nValProAspTyrPheHisLeuLeuMetGlyThrArgilleLeuPheLysAsnThrle 342  
Db 448 CATTCCTGACTGTGTGATCATCTCTGATGGGAGGAGAAATCTCTTCAAAATACATT 507  
Qy 342 uGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisAr 362  
Db 508 GGAACGTACACGGACTATTACTTGCATTATAAGTTAGAACAGCTATGTGAGAGCACCG 567  
Qy 362 gLeuValSerlleThrLeuLeuArgAspAlallePheCysGluAsnThrCluProAr 382  
Db 568 TTTGGTTTCTCTAATACCTGCTGAGAGATGCTGTGTCTGTGAGAAATCTGAAACCGG 627  
Qy 382 qSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyril 402  
Db 628 CTCTGTCCAGACACAGCAGAGAGCAGCAAGCAGACTTTCGAAGAAATGATGATACAT 587  
Qy 402 eProAspLeuLeuValLysCyslleGlyGluGluThrLysTyrGluSerilleArgLeuLe 422  
Db 688 TCCAGATTTAATAGGCAAGTGAT-GGGGAAGAGGCTTACATACGAGGATCAGGCTTCT 746  
Qy 422 uPheAsp 424  
Db 747 GTTTGAT 753

RESULT 15

BE874550  
LOCUS  
DEFINITION

BE874550 1076 bp mRNA linear EST 20-OCT-2000  
601489664F1 NIH\_MGC\_69 Homo sapiens cDNA clone IMAGE:3892095 5',  
mRNA sequence.

ACCESSION BE874550.1 GI:10323326

VERSION BE874550  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1076)

REFERENCE NIH-MGC http://mgs.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

Tissue Procurement: DCTD/DTF/Gazdar

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNase Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LAM9677 row: p column: 16

High quality sequence stop: 705.

Location/Qualifiers

## FEATURES

source

1..1076

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:3892095"

/tissue\_type="large cell carcinoma, undifferentiated"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 69"

/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.1 Kb. Library constructed by Life

Technologies."

BASE COUNT 388 a 164 c 233 g 291 t

## ORIGIN

## Alignment Scores:

Pred. No.: 7.12e-115 Length: 1076

Score: 1151.50 Matches: 243

Percent Similarity: 92.91% Conservative: 6

Best Local Similarity: 90.67% Mismatches: 14

Query Match: 47.25% Indels: 9

DB: 10 Gaps: 2

US-09-744-313A-1 (1-465) x BE874550 (1-1076)

Qy 79 ArgAsnLeuAlaAlaTTPlysIleSerIleProTyrValAspPheGluAspProSer 98

Db 2 CGAAACCTTCGTGCTGCAAGAAATTTAGCAATTCATATGATAGACTTTTGTAGGATCCCTCC 61

Qy 99 SerGluArgLysGluLysLysGluAqIleProValPheCysIleAspValGluArgAsn 118

Db 62 TCTGAAG 121

Qy 119 AspArgAlaValAlGlyHisGluProGluHisTyrSerValTyrArgArgTyrLeuGlu 138

Db 122 CATAG 181

Qy 139 PheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeu 158

Db 182 TTCTATGACTTGAATCAAACTAACAGAAATTTTCATGTCATTTCTCTGATGCCAGCTT 241

Qy 159 ProSerLysArgIleGlyProLysAsnTyrGluPheLeuLysSerLysAspGluGlu 178

Db 242 CCTTCTAAGAGAGATTCATGGCCCCCAAAATTTATGAATTTTAAAGTCAAGAGGAGAG 301

Qy 179 PheGlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeu 198

```
Db      ||| 302 TTCCAAGAAATATCTACAGAACTTCTGCAGCATCCAGAACTGAGTAATAGTCAACTCTG 361
QY      ||| 199 AlaaspPheLeuSerProAsnGlyGluThrGlnPheLeuAspLysIleLeuProAsp 218
Db      ||| 362 GCAGACTTTCTTCCCTTAATGGTGGGAAACACAAATTTCTTGATAAGATATCTACAGAT 421
QY      ||| 219 ValAsnLeuGlyLysIleLeuSerValProGlyLysLeuMetLysGluLysGlyGln 238
Db      ||| 422 GTAATCTTGGGAAATTAATAAATCTGTCTCGAAACTTAATGAAGAGAAAGGTCTAG 481
QY      ||| 239 HisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysPro 258
Db      ||| 482 CATTTGGAACTTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAGACA 541
QY      ||| 259 SerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysIleuPheAsn 278
Db      ||| 542 AGTAGACCAGAACTGACCAATTTCTCAG-CCTACTTCAGAAACACACAGAGCTTTTCAAT 600
QY      ||| 279 AspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluAsnGlyGlnAsn 298
Db      ||| 601 GATCTGTTTACAAATAATCAAAACCGTGTGA-AATACAGAGAGAAAGCAACATCAGAAAT 659
QY      ||| 299 TyrPheMet-GluValMet---ThrValGluGly-ValTyrAspTyrLeu-MetTyrVal 316
Db      ||| 660 TATTTTATCGAGGTGGATTGACTGTGTAGAGAGAGTCTATGATTACATGATGTATGTA 719
QY      ||| 317 GlyArgVal---ValPheGlnValProAspTrpLeuHisHisLeuLeuMetGlyThrArg 335
Db      ||| 720 AGGACGGGTAAATTTACCGAGTTCTCTGAATGGTTAAA-AATTTCTTAATGGGAAT-CGA 777
QY      ||| 336 IleLeuPheLysAsnThr 341
Db      ||| 778 AATCTCTTTAAAACACC 795
```

Search completed: January 31, 2004, 05:10:00  
Job time : 2958 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DOM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:43:12 ; Search time 5257 Seconds  
(without alignments)  
3618.599 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLHFCILIFRNTQKRGESF.....ELFPENKVKQKVTSTWSM 465

Scoring table: BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Fgapop 6.0, Dgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481336 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool/US09744313/runat.30012004.113535.4184/app.query.fasta\_1.647  
-DS=GenEmbl -QFMT=fastp -SUFFIX=rg -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 @CNG 1.1 3508 @runat.30012004.113535.4184 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_man.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	2437	100.0	1992	6	AX054819 Sequence
2	2387	97.9	1782	10	BC043328 Mus muscu
3	2382	97.7	3145	6	AX512835 Sequence
4	2382	97.7	3145	9	AK000362 Homo sapi
5	2380	97.7	2661	9	AY044865 Homo sapi
6	2380	97.7	3038	9	BC005110 Homo sapi
7	2371	97.3	2925	9	BC046520 Homo sapi
8	2358.5	96.8	3576	9	AK026479 Homo sapi
9	2354	96.6	3616	9	AK095380 Homo sapi
10	1964	80.6	1593	9	AF121863 Homo sapi
11	847	34.8	968	9	AF121863 Homo sapi
12	743	30.5	425	6	BD113855 EST and e
13	634	26.0	358	6	BD026500 Sequence
14	581	23.8	278898	2	AC116713 Mus muscu
15	578	23.7	129010	9	AL589666 Human DNA
16	447	18.3	224551	2	AC111832 Rattus no
17	447	18.3	278375	2	AC130093 Rattus no
18	442.5	18.2	183972	2	EX537259 Danio rer
19	314.5	12.9	84705	3	CEY48E1B Caenorhabdi
20	262	10.8	1617	9	AY044656 Homo sapi
21	262	10.8	2720	9	AK095661 Homo sapi
22	249	10.2	454	11	G30543 human STS S
23	229	9.4	1903	9	AK026814 Homo sapi
24	200.5	8.2	1964	9	BC029868 Homo sapi
25	198.5	8.1	2979	9	AF395843 Homo sapi
26	198.5	8.1	6049	9	D87443 Human mRNA
27	192.5	7.9	3040	10	BC004635 Mus muscu
28	192.5	7.9	5760	10	AK122231 Mus muscu
29	188	7.7	2955	6	BD160169 Primer fo
30	188	7.7	2955	9	AK023740 Homo sapi
31	188	7.7	3474	9	AF420470 Homo sapi
32	188	7.7	4129	9	BC045667 Homo sapi
33	184	7.6	1892	9	AK026643 Homo sapi
34	184	7.6	3832	9	BC041844 Homo sapi
35	182.5	7.5	4036	3	AY089224 Drosophila
36	182	7.5	1701	9	AF121862 Homo sapi
37	181	7.4	1472	3	AK117000 Ciona int
38	173	7.1	4879	10	AK122352 Mus muscu
39	168	6.9	34567	8	AF109957 S.pombe c
40	163.5	6.7	1803	10	AF312007 Mus muscu
41	156	6.4	1699	9	AK026637 Homo sapi
42	156	6.4	2065	9	BC005022 Homo sapi
43	156	6.4	2271	9	AF121859 Homo sapi
44	156	6.4	2310	9	AF131214 Homo sapi
45	156	6.4	4175	9	AF172847 Homo sapi

# ALIGNMENTS

RESULT 1

AX054819  
LOCUS AX054819 1992 bp DNA linear PAT 13-JAN-2001  
DEFINITION Sequence 3 from Patent WO0073334.  
ACCESSION AX054819  
VERSION AX054819.1 GI:12228268  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.  
REFERENCE 1  
AUTHORS Yue, H., Tang, Y. T. and Azimzai, Y.  
TITLE Human sorting nexins  
JOURNAL Patent: WO 0073334-A 3 07-DEC-2000;  
Incyte Genomics, Inc. (US)  
FEATURES  
source  
1..1992  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 671 a 308 c 372 g 641 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.5e-203 Length: 1992  
Score: 2437.00 Matches: 465  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-09-744-313A-1 (1-465) x AX054819 (1-1992)  
QY 1 MetTyrLeuIleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPhe 20  
Db 116 ATGTACTTGATACATTTTGTGTTTAAATTCAGAGACACACAGAAAGGGGAGAAATCATTT 175  
QY 21 GlyIleSerArgIleGlySerLysIleGlyValPheLysSerThrThrMetGluGly 40  
Db 176 GGNATCAGCAGATAGTACCAAAATTAAGGNGATTCACAAAGTACCACATGGAGGA 235  
QY 41 AlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGlyIle 60  
Db 236 GCTATGTGCTAAATATGCTGAGTGAAGTGAAGATGATTTTATTGAGAGGTATT 295  
QY 61 ValValMetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80  
Db 296 GTTGTAAATGAGATGATTCCTCAGTGGAGGCTGTGAGCACACCTAATACTCCCGAAAC 355  
QY 81 LeuAlaIleTyrLysIleSerIleProTyrValAspPheGluAspProSerSerGlu 100  
Db 356 CTTCGTCATGGAATAATAGCATTCATATGATGATGATGATGATGATGATGATGATGAT 415  
QY 101 ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120  
Db 416 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 475  
QY 121 ArgAlaValGlyHisGluProGluHisTyrSerValTyrArgArgTyrLeuGluPheTyr 140  
Db 476 AGAGCAGTTCGACACGAGCCTGAACATGCTGCTCTATAGAAGATATCTGAATCTAT 535  
QY 141 ValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSer 160  
Db 536 GTACTTGAATCAAACTAACAGAAATTCATGCGGATTCCTGATCCCGAGTTCCTCT 595  
QY 161 LysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGln 180  
Db 596 AAGAGGATATTCGCCCCCAAAATTAATGAATTCCTTAAGTCAAGAGGAGGAGGAGGAGGAG 655  
QY 181 GluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
Db 656 GAATATCTACAGAACTTCTGCAGCATCCAGAACTGAGTAATAGTCACTTCTGCGAGAC 715

QY 201 PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220  
Db 716 TTTCTTTCCCTTAATCGTGGGAAACACAAATTTCTTGATAAGATACCTACAGATGAAAT 775  
QY 221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu 240  
Db 776 CTTCGGAGAAATATAAAATCTGTTCTCGAAACCTAATGAAGAGAAAGTTCAGCATTTG 835  
QY 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
Db 836 GAACCTTTTATCATGAATTTTCAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGA 895  
QY 261 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnLeu 280  
Db 896 CCAGAACTGACCAATCTCAGCCCTACTTCAGAAACCAACCAAGAGCTTTTCAATGATCTG 955  
QY 281 PheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPhe 300  
Db 956 TTTTAAATAATATGCAACCCGCTGCTGAAATACAGAGAGAAAGCAAAATCAGAAATATTT 1015  
QY 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
Db 1016 ATGGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGTATGAGGAGGTAGTT 1075  
QY 321 PheGlnValProAspTyrPheHisLeuLeuMetGlyThrArgIleLeuPheLysAsn 340  
Db 1076 TTCAGGTTCCCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAATCCTCTTTAAAC 1135  
QY 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu 360  
Db 1136 ACCCTGGAATGATATCTACTGATTTACTATCTTCAGTGTAACTAGAACAGCTATTTT 1195  
QY 361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380  
Db 1196 CACGTTGTGCTCATCTAATACACATCTCAGAGATGCTATATCTCTGTGAACACTGAA 1255  
QY 381 ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400  
Db 1256 CCTCGCTCTCTCCAGATAGCAAGCAAAAGAGGCAAAACAGACTTTTGAAGAAATGATGAAT 1315  
QY 401 TyrIleProAspLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArg 420  
Db 1316 TACATTCAGATCTGTGTAGTCAAGTGTATTGGTGAAGAAACCAAGATGAAAGCATCAGA 1375  
QY 421 LeuLeuPheAspGlyLeuGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
Db 1376 CTTCCTGTTGATGCTTACAGCAACCAAGTACTCAGACGAGCTGACTTATGTTTATTG 1435  
QY 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
Db 1436 GACATCTGTATACAGGAACCTGTTTCCAGAGCTCAATAAGGTACAAAGGAAGTTACCTCT 1495  
QY 461 ValThrSerTyrMet 465  
Db 1496 GTGACATCTTGATG 1510  
RESULT 2  
BC043328 1782 bp mRNA linear ROD 16-APR-2003  
LOCUS Mus musculus RIKEN cDNA C330035N22 gene, mRNA (cDNA clone MGC:49424  
DEFINITION IMAGE:5400157), complete cds.  
ACCESSION BC043328  
VERSION BC043328.1 GI:27694048  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1782)  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

TITLE	Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Carrinci, P., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Scarinci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	
	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903 (2002)
MEDLINE	22388257	
PUBMED	12477932	
REFERENCE	2 (bases 1 to 1782)	
AUTHORS	Straussberg, R.	
TITLE	Direct Submission	
JOURNAL	Submitted (09-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>	
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs@mail.nih.gov">cgabs@mail.nih.gov</a> Tissue Procurement: The Cepko Laboratory CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Boedet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.	
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAC Plate: 86 Row: P Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.	
	Location/Qualifiers 1. .1782 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="MGC:49424 IMAGE:5400157" /tissue_type="Eye, retina, mouse strain C57Bl/6" /clone_lib="NIH_MGC_94" /lab_host="DH10B" /notes="Vector: pCMV-SPORT6" 1. .1782 /gene="C330035N22R1k" /notes="synonym: B830022K16" /db_xref="LocusID:244962" /db_xref="MGI:2443156" 228. .1616 /codon_start=1 /product="C330035N22R1k protein" /protein_id="AA043328.1" /db_xref="GI:27694049" /db_xref="LocusID:244962" /translations="MRICLIFRSTQKRGSGFISRGISKGVFKSTTWEGAVLPNYG VAEGEDDFIEGIVMEDDSPVEAVSTPNTPRNLAANKISIPVDFDFPSERKEKK	
gene		
CDS		

BASE COUNT		567 a	339 c	380 g	496 t
ORIGIN		VQKEATSMTSMW			
Alignment Scores:		1.1e-198			
Pred. No.:		2387.00			
Score:		2387.00			
Percent Similarity:		99.14%			
Best Local Similarity:		97.42%			
Query Match:		97.95%			
Indels:		0			
Gaps:		10			
US-09-744-313A-1 (1-465) x BC043328 (1-1782)					
Qy	1	MetTyrIleuIleHisPheCysLeuIlePheAtrGAsnThrGlnLysArgGlyGluSerPhe	20		
Db	219	ATGTACTGATGCGTATTGTTTAAATTTAGAGCACAATAAGAGAGAGAGATGTTT	278		
Qy	21	GlyIleSerArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGly	40		
Db	279	GGATCAGCAGATAGTAGTACCAAAATTAAGGCGTATTCAAGAGTAGCACAATGAGGGA	338		
Qy	41	AlaMetIeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGlyIle	60		
Db	339	GCTGTGCTGCTTAATTACGGGTGGCTGAAGCGAAGATGACTTTATTGAAGAAGGATT	398		
Qy	61	ValValMetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn	80		
Db	399	GTGGTAATGGAGGATGACTCTCAGTAGAGCTGTGAGCACACTTAACCTCTCTGAAAC	458		
Qy	81	LeuAlaAlaTrpLysIleSerIleProTyrValAspPheGluAspProSerSerGlu	100		
Db	459	CTTGCTGCTGGAAAATTAGCATTGCTGATGAGACTTTTTCAGAGATCTCTCTCGAA	518		
Qy	101	ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg	120		
Db	519	AGGAAGAAGAAGAGAGAGAGATTCTGTGTTTGTATCGATTGTAAGAAGACGACAGA	578		
Qy	121	ArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyr	140		
Db	579	CGAGCAGTTGCACACGAGCTGAGCATTGCTGCTCTATAGAAGATATCTTGAATCTAT	638		
Qy	141	ValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSer	160		
Db	639	GTTCTTGATCAAACTAACAGAAATTTTCATGGGCACATTTCTCTGATGCTCAGTCTCC	698		
Qy	161	LysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGln	180		
Db	699	AAAAGGATCATTTGCCCCAAAATACTAGTATTTTGAAGTCGAGAGAGAGAGATTCAG	758		
Qy	181	GluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp	200		
Db	759	GAGTATCTGCAGAACTTGTGCACACCGGAGCTGAGTAACAGCAGCTCTGGCTGAC	818		
Qy	201	PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn	220		
Db	819	TTTCTCTCCCCAACGGTGGGAGACACAGTTTCTAGATAAGATACTTCCAGATGTAAAT	878		
Qy	221	LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu	240		
Db	879	CTTGGAAAATAATAAGTCTGTTCTCCGGGAACTAATGAAGAGAAAGGTCAGATTG	938		
Qy	241	GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg	260		
Db	939	GAGCATTTCATCAGTCTTCAATTAATCTTGGGAATCTCCCAACCTAAACCGAGTCGG	998		
Qy	261	ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu	280		



999 CCAGAACTGACGATTTCTAGCCCACTTCCGAAAACCAAGAACTTTTCAAGATCTG 1058  
281 PheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPhe 300  
1059 TTTAAGAAATATCAAAACCGCTGAGAACACAGAGCGAAGCAAAATCAGAACTACTTC 1118  
301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
1119 ATGAGGTGATGACTGTGACGAGCTATGATTAACCTGATGATGATGAGCGAGTAGTT 1178  
321 PheGlnValProAspTyrPheHisLeuLeuMetGlyThrArgLysLeuPheLysAsn 340  
1179 TTTCAAGTCCAGACGGCTTCATCTCTTAATGGGAACCGAATCTCTTTAAGAAC 1238  
341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu 360  
1239 ACCCTGGAAATGACACAGACTACTACCTCCAGTGCAGCTGAGGAGCTGTTTCAAGG 1298  
361 HisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaLeuPheCysGluAsnThrGlu 380  
1299 CACCGGTGCTCTCCCTCATCACCCTCTCAGAGATGCTATATTTTGTGAAAATACTGAA 1358  
381 ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400  
1359 CCTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAAT 1418  
401 TyrIleProAspLeuLeuValLysCysLysGlyGluThrLysTyrGluSerIleArg 420  
1419 TACATTCAGATCTGATAGTCAAGTGTATCGGTGAAGAACCAAGTACCAAGACATCAGA 1478  
421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
1479 CTCCTGTTGATGCGCTGACAGCCAGCTGCTCAACAGAGCTGACCTATGTTGCTG 1538  
441 AspIleValIleGlnLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
1539 GACATTGTGATACAGGAGCTGTTTCCAGAGCTAAATAAGGTACAAAAGGAGCTACCTCC 1598  
461 ValThrSerTyrMet 465  
1599 ATGACATCCTGGATG 1613

RESULT 3  
AX512835 3145 bp DNA linear PAT 03-OCT-2002  
LOCUS Sequence 4 from Patent WO02062839.  
DEFINITION AX512835  
ACCESSION AX512835  
VERSION AX512835.1 GI:23504019  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Daenen,M.J., Cleutjens C.B. and Zaman,G.J.  
TITLES Markers of unstable atherosclerotic plaques  
JOURNAL Patent: WO 02062839-A 4 15-AUG-2002;  
Universiteit Maastricht (NL)  
FEATURES  
source  
1. 3145  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 985 a 570 c 659 g 931 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6,03e-198 Length: 3145  
Score: 2382.00 Matches: 454  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.56% Mismatches: 0  
Query Match: 97.74% Indels: 0

DB: 6 Gaps: 0  
US-09-744-313A-1 (1-465) x AX512835 (1-3145)  
Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIle 29  
Db 1598 TTTCCGAAACACACAGAAAGGGGAGAAATCATTTGGAAATCAGCAGAAATAGTAGCAAAAT 1657  
Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
Db 1658 AAAGGAGTATTTCAGAAAGTACCAACATGGAGGGAGCTATGTTGCCTAATATATGTTAGCT 1717  
Qy 50 GluGlyValAspPheIleGluGlyValValMetGluAspSerProVal 69  
Db 1718 GAAGGTGAAGATGATTTTATTAAGAGGATTTGTTGTAATGGAAGATGATTTCCAGTG 1777  
Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
Db 1778 GAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGCTGCTGCAATAGCAATTTAGCAATCCA 1837  
Qy 90 TyrValAspPhePheGluAspProSerSerGluArgLysGlnLysLysGluArgIlePro 109  
Db 1838 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1897  
Qy 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 1898 GTGTTTGTATGTATGTGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCTGACAT 1957  
Qy 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 1958 TGGTCTGTCTATAGAGATATCTGAATTTCTATGACTTGAATCAAAATCAAGAAATTT 2017  
Qy 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 2018 CATGTGCTATTTCTGATGCCAGCTTCCCTTCTAAGAGAGATCATTTGGCCCCCAAAATAT 2077  
Qy 170 GluPheLeuLysSerLysArgGluGluPheGlnGlyTyrLeuGlnLysLeuGluHis 189  
Db 2078 GAATTTCTTAAAGTCAAGAGAGGAGGAGTCCAGAAATATCTACAGAACTTCTGACGAT 2137  
Qy 190 ProGluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 2138 CCGAACTGAGTATATAGTCACTTCTGGAGACTTTCTTTCCCTTAATGGTGGGGAACA 2197  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
Db 2198 CAATTTCTTGATAGATAGTACTACAGATGTAATCTTGGGAAAATATATAAAATCTGTTCCT 2257  
Qy 230 GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
Db 2258 GGAATAACTAATGAAAGAGAGAGGTCAGCATTTGGAACCTTTTATCATGAATTTTCAAT 2317  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 2318 TCTTTGTAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACT 2377  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 2378 TCAGAAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 2437  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 2438 AATACAG 2497  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis 329  
Db 2498 TATGATTACCTGATGTATGTAGGACGGGTAGTTTTCAGATTTCTTCAGATTCCTGCTCAT 2557  
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 2558 CTCCTTAATGGAACTCGAATCTCTTTTAAACACCTCGAAATGTAATGATGATGATGAT 2617  
Qy 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369

Db 2618 CTTGAGTGAACATAGAACACAGCTATTTTCAGAGACACCGTTGGTCTCCTCATTAACACTT 2677  
Qy 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
Db 2678 CTCAGAGATGCTATATCTTGTGAAACACATGACCTCGCTCTCTCCAGATAGCAAAAA 2737  
Qy 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
Db 2738 GGAGCAAAACAGACTTTTGAAGAAATGATGAATACATTCAGATCTGTAGTCAAGTGT 2797  
Qy 410 IledGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
Db 2798 ATTGGTGGAGAAACCAAGATATGAAGCATCAGACTTCTGTTTGAATGGCTTACAGCAACCA 2857  
Qy 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
Db 2858 GTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATACAGGACTGTTCCTCA 2917  
Qy 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
Db 2918 GAGCTCAATAGGTACAAAGAGGTACCTCTGTGACATCTGGATG 2965  
AK000362 3145 bp mRNA linear PRI 22-FEB-2000  
Homo sapiens cDNA FLJ20355 fis, clone HEP15804, highly similar to  
AF121863 Homo sapiens sorting nexin 14.  
AK000362  
AK000362.1 GI:7020397  
oligo capping; fis (full insert sequence).  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,  
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,  
Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Nakamura, Y., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
Unpublished  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
TITLE  
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,  
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,  
Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source 1. .3145 Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP15804"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_lib="HEP"  
/notes="cloning vector pME18SFL3"  
misc\_feature 1. .3145  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14"  
BASE COUNT 985 a 570 c 659 g 931 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.03e-198 Length: 3145  
Score: 2382.00 Matches: 454  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.58% Mismatches: 0  
Query Match: 97.74% Indels: 0  
DB: 9 Gaps: 0  
US-09-744-313a-1 (1-465) x AK000362 (1-3145)  
Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
Db 1598 TTTTCGAAACACACAGAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGTAGCAAAAT 1657  
Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
Db 1658 AAAGGAGTATTGAGAAAGTACCACAAATGGAGGAGCTATGTTGCCCTAAATATGGTGTAGCT 1717  
Qy 50 GluGlyGluAspAspPheIleGluGluGlyIleValMetGluAspAspSerProVal 69  
Db 1718 GAAGGTGAAGATGATTTTATTCAGAGAGGTATTTGTTATGAGAGATGATTTCTCCAGTG 1777  
Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
Db 1778 GAGGCTGTGAGCACACCTAATACTCCCGAAACCTTGCTGCATGGAAAATATAGCATTTCA 1837  
Qy 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109  
Db 1838 TATGTAGACTTTTTCAGGATCCCTCTCTGAAGGAGGAGGAGAAAAGAAAGAAATTCCT 1897  
Qy 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 1898 GTGTTTGTATTGATGTTGAAGAAATGATAGAGAGAGCTGGACACAGGCTGACAT 1957  
Qy 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 1958 TGGTCTGCTATAGAGAGATATCTTGAATCTTATGTAATCAAACTAAACAGAAATTT 2017  
Qy 150 HisGlyValPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 2018 CATGGTGCAATTCCTGATGCCAGCTTCCTCTTAAGAGGATCATTTGGCCCCCAAAATAT 2077  
Qy 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
Db 2078 GAATTTCTTAAAGTCAAGAGGAGAGAGTTCGAAGAATATCTACAGAACTTCGCGAGCAT 2137  
Qy 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 2138 CCAGAACTGAGTAAATAGTCAACTTCCTGGCAGACTTTCTTCCCTTAATGGTGGGAAACA 2197  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
Db 2198 CAATTTCTTGAAGATACCTACAGATGTAAATCTTTGGGAAAATATATAAATCTGTTCCT 2257  
Qy 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheLeuAsn 249  
Db 2258 GGAATACTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAAATTTCAATTAAT 2317  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 2318 TCTTGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGTACCACTTCTCAGCCCTACT 2377  
Qy 270 SerGluAsnLeuLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 2378 TCAGAAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 2437  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 2438 AATACAGAGAGAAAGCAAAATCAGAAATTAATTTATGAGGAGGTGATGATCTGTAGAGGAGTC 2497  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHis 329  
Db 2498 TATGATTACCTGATGTATGTAGGACGGGTAGTTTTCAGATTTCTGACTGCTTCATCAT 2557

BASE COUNT 855 a 490 c 534 g 782 t

ORIGIN

Alignment Scores:

Pred. No.: 7 34e-198 Length: 2661

Score: 2380.00 Matches: 455

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 97.66% Indels: 0

DB: 9 Gaps: 0

US-09-744-313a-1 (1-465) x AY044865 (1-2661)

QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgIleGlySerLysLys 30

DB 1294 AGGAACACACAGAAAAGGGAGAAATCATTTGCAATCAGCAGAAATAGGTAGCAAAATTTAA 1353

QY 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50

DB 1354 GGAGTATTCAAAAGTACCACAAATGGAGGAGCTATGTCCTTAATATGCTAGTGTGAA 1413

QY 51 GlyLysAspPheIleGluGluGlyIleValValMetGluAspAspSerProValGlu 70

DB 1414 GGTGAAGATGATTTTATGAAGAGGTATGTTGTATGGAAGATGATTTCTCCAGTGGAG 1473

QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyr 90

DB 1474 GCTGTGAGCACACTAATCTCCCGAACCTTGTCTGCATGGAAAATTTAGCATTTCCATAT 1533

QY 91 ValAspPheGluAspProSerSerGluArgLysGluLysLysLysLysLysLysLys 110

DB 1534 GTAGACTTTTTCAGGATCCCTCTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1593

QY 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHISgluProGluHisTrp 130

DB 1594 TTTTGTATGATGTTGAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1653

QY 131 SerValTyrArgTyrLeuGluPheTyrValLeuGluSerLysLysLysLysLysLys 150

DB 1654 TCTGTCTATAGAGATATCTTGAATCTATGATCTTGAATCAAACTAACAGAAATTTAT 1713

QY 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170

DB 1714 GGTGCAATTTCTGATGCCAGCTTCTCTTCTAGAGGATCATTTGCCGCCCAAAATTTAT 1773

QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLysLysLysLys 190

DB 1774 TTTCTAAAGTCAAG 1833

QY 191 GluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210

DB 1834 GAATGAGTAAATAGTCAACTTCTGGCAGACTTTTCTTCTTCTTCTTCTTCTTCTTCT 1893

QY 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValProGly 230

DB 1894 TTTCTTGTATAGATACCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCTCT 1953

QY 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheLeuAsnSer 250

DB 1954 AAACCTAAAGAAAG 2013

QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270

DB 2014 TGTGAGTCTCAAG 2073

QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290

QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349

DB 2558 CTTCTTAATGGGAACTCGAATCTCTTTTAAACACACCTCGGAAATGATATGATATAT 2617

QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369

DB 2618 CTTCTAGTAACTAGAACAGACTATTTTCAGAGACACCGTTTGGTCTCACTATAACACT 2677

QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389

DB 2678 CTTCTAGATGCTATATCTCTGTAACACACTGAACTCGCTCTCTCCAGATTAACAAAA 2737

QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuValLysCys 409

DB 2738 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTCATCCAGATCTGTTAGTCAAGTGT 2797

QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429

DB 2798 ATTGTGAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGAAGCTTACAGCAACCA 2857

QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449

DB 2858 GTACTCAACAGAGCTGACTTATGTTTATTGGACATTTGATACAGAGAACTGTTTCCA 2917

QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465

DB 2918 GAGCTCAATAAGGTACAAAGGAAGTTACCTCTGTGACATCTTGGATG 2965

RESULT 5

AY044865

LOCUS

DEFINITION

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

Hom sapiens (human)

Hom sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2661)

Hong, W.

The complete coding region of SNX14

Unpublished

REFERENCE

2 (bases 1 to 2661)

Hong, W.

Direct Submission

Submitted (11-JUL-2001) Membrane Biology Laboratory, Institute of

Mol & Cell Biol, 30 Medical Drive, Singapore 117609

Location/Qualifiers

1. .2661

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

1. .2661

/gene="SNX14"

/gene="SNX14"

/note="contains PX domain"

/codon\_start=1

/product="sorting nexin 14"

/protein\_id="AAK97796.1"

/db\_xref="GI:15529064"

/translation="MSQKLQRLLDVGRICRQYPLFCFLLLCLSAALLNRYVHI

LMFWFVAGVWTFYCSLGDPSLLPNIPFTIKYKPKGLGLFPGHSCAVGKVC

KRHPSPILLNENYQFWDLLKISSKVDASLSEVDIPSLITKKLKAAMKHLEIVYKQK

VKNTEFTQQALAEYGEPLHVALSRDELHRLKLTLLFPYLPKPKATDCRSLLL

IRETLGSVPLSPDLFDPTVNHLLIIPIDPSPEKATEPASPPLVPLQKFAEPN

KPSVLKLELQREOQLLRFWMFLKQGAHVLFQCLTVEEFNDRILRPELSENDE

MLSHLELQTYKTYCLDESIDKTRDPFIVERIORIAEGPYIDVVKLOTMRCLPEAY

EHVLSLENVTFMPCFCHSDRYFQLRLGASPTNSKLNENKQKCESGIESKSI

KGVKFTMTSGAMLPNTGVAGEDDFIEEGIVMDDSPVEAVSTNTNPNLAAWKIS

IPYVDFPFSSSEKKEKKEIPVCFIDVNRNDRVAGHPEHWSVYRYLEFVLESK

2074	GAAAACAACAGNAGCTTTTCAATGATCTGTGTTTTAAAAATAATGTCAAACCGTGCTGGAAT	2133
QY	ThrGluArgLysGlnAsnGlnAsnIlePheMetGluValMetThrValGluGlyValTyr	310
Dd	2134 ACAGAGAGAAGAACAAATCAGAAATTATTTATGGAGGTGATGACTGTAGAAGGAGTCTAT	2193
QY	AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpIleHisHISLeu	330
Dd	2194 GATTACCTCATGTATGTAGGACGGTAGTFTTCCAGGGTTCCCTGACTGGCTTCATCATCTC	2253
QY	LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu	350
Dd	2254 TTAAATGGGAACCTCGAATCCCTCTTTTAAAAACACCCTCGGAATGTATACTGNATTACTATCTT	2313
QY	GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeulleThrLeuLeu	370
Dd	2314 CAGTGTAAACTAGACAGACTATTTCAGGASCACCGTTTGGTCTCACTCATTAACACTTCTC	2373
QY	ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly	390
Dd	2374 AGAGATGCTATATTCGTGTAACACACTGAAACCTCGCTCTCTCCAAGATAAGCAAAAGGA	2433
QY	AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysVelle	410
Dd	2434 GC AAAACAGACTTTTGTGAAGAAATGATGAATTCACATCCAGATCTGTGTAGTCAAGTGTATT	2493
QY	GlyGluGluThrLysTyrGluSerILEArgLeuLeuPheAspGlyLeuGlnGlnProVal	430
Dd	2494 GGTTGAAGAAACCAAGTATGAAGCAATCAGACTTCTGTTTGTGTCCTTACAGCACCAAGTA	2553
QY	LeuAsnLysGlnLeuThrTyrValLeuLeuAspileValIleGlnGluLeuPheProGlu	450
Dd	2554 CTCACACAGCAGCTGACTTATGTATTTTATGGACATTTGTATACAGGAACCTGTTTCCAGAG	2613
QY	LeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet	465
Dd	2614 CTCAATTAAGGTACAAAAGGAAGTACTCTCTGTGACATCTTTGGATG	2658

RESULT 6	
BC005110	3038 bp mRNA linear, PRI 12-JUL-2001
LOCUS	BC005110
DEFINITION	Human sapiens, clone MGC:13217 IMAGE:3959086, mRNA, complete cds.
ACCESSION	BC005110
VERSION	BC005110.1 GI:13477272
KEYWORDS	MGC.
SOURCE	Human sapiens (human)
ORGANISM	Human sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 3038)
JOURNAL	Strausberg, R.
	Direct Submission
	Submitted (26-MAR-2001) National Institutes of Health, Mammalian
	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk
	Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a>
	Tissue Procurement: ATCC
	cDNA Library Preparation: Rubin Laboratory
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Genome Sequence Centre,
	BC Cancer Agency, Vancouver, BC, Canada
	<a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a>
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
	Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
	Ness, Pawan Pandoh, Anna-Tilisa Prabhu, Parvaneh Saedi, Jacqueline
	Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
	George Yang, Scott Zuyderduyn, Marco Marra.

1782 TTTTGTATTGATTTGAAAGAAATGATAGACAGCAGTGGACACGACCTGGAACATTGG 1841  
 QY 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
 Db 1842 TCTGTCTATAGAGATATCTTGAATCTATGACTTGAATCAAACTAACAGATTTCAT 1901  
 QY 151 GluValaPheProAspAlaGlnLeuProSerLysArgIleLeuGlyProLysAsnTyrGlu 170  
 Db 1902 GTTGCAATTCCTGATGCCAGCTTCTTCTAAGAGATCATTTGGCCCCCAAAATATGAA 1961  
 QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
 Db 1962 TTCTTAAATCAAGAGGGAAGATTTCCAAGATATCTACAGAACTTCTTCGAGATCCA 2021  
 QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
 Db 2022 GAACTGAGTAATAGTCAACTTCGCGAGACTTTTCCCTTAATGTTGGGGAACACAA 2081  
 QY 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValProGly 230  
 Db 2082 TTCTTGTATAGATATACACAGATGTAATCTTGGGAAATATATAAATCTGTTCTCGA 2141  
 QY 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheLysLeuMetAsnPheLysSer 250  
 Db 2142 AACTAATGAAGAGAAAGGTGACATTTGGACCTTTTATCATGAATTCATTAATCT 2201  
 QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
 Db 2202 TGTGAGTCTCCAAAGCCCTAAACCAAGTAGACCACTGACCAATCTCAGCCCTACTTCA 2261  
 QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290  
 Db 2262 GAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAT 2321  
 QY 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
 Db 2322 ACAGAGAGAAAGCAAAATCAGATTAATTTATGGAGGTGATGACTGTAGAAGGAGTCTAT 2381  
 QY 311 AspTyrLeuMetTyrValGlyArgValaPheGlnValProAspTrpLeuHisLeu 330  
 Db 2382 GATTACCTGATGATGATAGGACGGGTAGTTTTCCAGGTCTCTGACTGGCTTCATCATCTC 2441  
 QY 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
 Db 2442 TTAATGGGAACCTCGAATCTCTTTAAACACCCCTGGAAATGATATCTGATTACTATT 2501  
 QY 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
 Db 2502 CAGTGTAACTAGACAGCTATTTCCAGGACCCGTTTGGTCTCACTCATACACACTCTC 2561  
 QY 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
 Db 2562 AGAGATGCTATATTCTGTGAAACACACTGAACCTGCTCTCTCCAGATAAGCAAAAGGA 2621  
 QY 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
 Db 2622 GCAAAACAGACTTTTGAAGAAATGATGATTAATTCAGATCTGTTAGTCAAGTGATT 2681  
 QY 411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
 Db 2682 GTGGAAGAACCAAGATATGAAGCATCAGACTCTCTGTTGATGGCTTACACCAACAGTA 2741  
 QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
 Db 2742 CTCACAGACAGCTGACTTATGTTTATTTGGACATTGTGATACAGAACTGTTTCCAGAG 2801  
 QY 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
 Db 2802 CTCATAGGTATCAAAAGGAAGTACCTCTGTGACATCTTGGATG 2846

RESULT 7

BC046520

LOCUS

2925 bp

mRNA

linear

PRI 21-FEB-2003

## DEFINITION

Homo sapiens, Similar to sorting nexin 14, clone IMAGE:5267454,  
 mRNA.

## ACCESSION

BC046520  
 VERSION  
 BC046520.1 GI:28461364

## KEYWORDS

Source

## ORGANISM

Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 2925)

## AUTHORS

Strausberg, R.

## TITLE

Direct Submission

## JOURNAL

Submitted (03-FEB-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk  
 Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegod, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: TRAK Plate: 91 Row: a Column: 6.

## FEATURES

source

1..2925  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5267454"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH MGC 97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

BASE COUNT 948 a 513 c 579 g 885 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,02e-197 Length: 2925  
 Score: 2371.00 Matches: 454  
 Percent Similarity: 99.7% Conservative: 0  
 Best Local Similarity: 99.7% Mismatches: 1  
 Query Match: 97.2% Indels: 0  
 DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x BC046520 (1-2925)

## QY

11 ArgAsnThrGlnLysArgGlyCysSerPheGlyIleSerArgIleGlySerLysIleLys 30

## Db

1381 AGGAACACACAGAAAAGGGAGAAATCTTGGATTCAGAGATAGGTAGCAAAATTTAA 1440

## QY

31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValaGlu 50

## Db

1441 GGAGTATTCAAAGATGATACCAATGAGGAGGTATGTTGCCTAATATGCTAGCTGAA 1500

## QY

51 GlyGluAspAspPheIleGluGlyIleValMetGluAspAspSerProValGlu 70

## Db

1501 GGTGAAGATGATTTTATTGAAGAAGGTATTTGTTGTAATGGAAGATGATTCTCCAGTGGAG 1560

## QY

71 AlaValSerThrProAsnThrProArgAsnLeuAlaIleTyrLysIleSerIleProTyr 90

## Db

1561 GCTGTGACACACCTAATATCTCCAGAAACCTTGCTGCTGCAATAGCAATTCATAT 1620

91 ValAspPheGluAspProSerGluArgLysGluLysLysGluArgIleProVal 110  
1621 GTAGACTTTTGGAGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTG 1680  
111 PheCysIleAspValGluArgAspArgAlaValGluHisGluProGluHisTyr 130  
1681 TTTTGTATTGATGTTGAAAGAAATGATAGGAGAGAGATTTGGACAGCCCTGGAACATTTGG 1740  
131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
1741 TCTGCTATAGAGATATCTTGATTCATGTACTTGAATCAAACTAACAGAAATTCAT 1800  
151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleGluProLysAsnTyrGlu 170  
1801 GGTGCAATTCCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGGCCCAAAATATATGAA 1860  
171 PheLeuLysSerLysArgGluGluPheGlnTyrLeuGluLysLeuGluHisPro 190  
1861 TTCTTAAGTCAAGAGGAGGAGATTTCCAGAAATATCTACAGAACTTCTTGAGCATCCA 1920  
191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGluGluThrGln 210  
1921 GAATGAGTAATAGTCACTTCTGGCAGACTTCTTCTCCCTAATGTTGGGGAACACAA 1980  
211 PheLeuAspLysIleLeuProAspValAsnLeuGluLysIleIleLysSerValProGly 230  
1981 TTCTTGAATAGATACCTACAGATGTAATCTTGGGAAATATATAAATCTGTTCTGCA 2040  
231 LysLeuMetLysGluLysGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
2041 AAATAATGAAGAGAAAGGTGACATTTTGGAACTTTTATCATGAATTTCAATTAATCT 2100  
251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
2101 TGTGAGTCTCAAGGCTTAACCAAGTAGACAGAACTGACCAATTTCTGAGCCCTACTCA 2160  
271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290  
2161 GAATACCAAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAAAT 2220  
291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
2221 ACAGAGAGAAAGCAAAATCAGATTTATTTATGAGGTGATGATGCTAGAGAGATCTAT 2280  
311 AspTyrLeuMetTyrValGlyArgValPheGlnValProAspTyrPheHisLeu 330  
2281 GATTACCTGATGATGTAGACGGGTATTTTCCAGGTTCTCTGACTGGCTTCATCATCTC 2340  
331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
2341 TTAATGGAACTCGAATCCCTCTTTAAACACACCTGGAAATGATATCTGATTAATCTT 2400  
351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
2401 CAGTGTAACTAGACAGCTATTTACAGAGCACCGTTTGTCTCACTCAATACATCTTC 2460  
371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
2461 AGAGATGCTATATTTCTGTGAAAACACCTGCACTCTCTCTCCAGATATAGCAAAAGGA 2520  
391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
2521 GCAAAACAGACTTTTGAAGAAATGATGAATTCGATTCGATCTGTTAGTCAAGTGTATT 2580  
411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVal 430  
2581 GGTGAAGAAACCAAGATGAAAGCATCAGACTTCTGTTTGTATGCTTACAGCACCGTA 2640  
431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
2641 CTCACAGAGAGCTGATTAATGTTTATTTGACATTTGTGACATTTGTATGACAGGAATCTTTCCAGAG 2700

451 LeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
2701 CTCATTAAGGTACAAAAGGAGGTACTCTCTGTGACATCTTGGATG 2745

RESULT 8  
AK026479 3576 bp mRNA linear PRI 29-SEP-2000  
Homo sapiens cDNA: FLJ22826 fis, clone KAI4022, highly similar to  
AF121863 Homo sapiens sorting nexin 14 (SNX14) mRNA.  
AK026479  
AK026479.1 GI:10439350  
oligo capping; fis (full insert sequence).  
Homo sapiens (human)  
Homo sapiens  
Sukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,  
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Tanigami, A.,  
Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M.,  
Amori, Y., Oka, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T.,  
Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 3576)  
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639,  
Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,  
Fax: 81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source  
1..3576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KAI4022"  
/tissue\_type="ileal mucosa"  
/clone\_lib="kaia"  
/note="cloning vector pME18SFL3"  
misc\_feature  
1..3576  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14 (SNX14) mRNA"  
BASE COUNT 1181 a 633 c 735 g 1027 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7,91e-196 Length: 3576  
Score: 2358.50 Matches: 453  
Percent Similarity: 99.13% Conservative: 1  
Best Local Similarity: 98.91% Mismatches: 1  
Query Match: 96.78% Indels: 3  
DB: 9 Gaps: 1  
US-09-744-313A-1 (1-465) x AK026479 (1-3576)  
QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLys 30  
Db 2017 AGGAACACACAGAAAAGGGAGAGATCATTTGGAAATCAGCAGATAGGTAGCAAAATATAA 2076  
QY 31 GlyValPheLysSerThrThrMetGluClyAlaMetLeuProAsnTyrGlyValAlaGlu 50  
Db 2077 GGAGTATTCAAAAGTACCACAAATGAGGAGCTATGTTGCCCTAATATTGGTGTAGCTGAA 2136

51 GlyIuAspPheIleGluGlyIleValMetGluAspSerProValGlu 70  
2137 GGTGAAGATGATTTTATTTGAAGAAGGATATTTGTTGAATGAAGATGATTTCCAGTGGAG 2196  
71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTyrLysIleSerIleProTyr 90  
2197 GCTGTGAGCACACCTTAATCTCCCGAAACCTTCTGTCATGGAATAATAGCATTCATAT 2256  
91 ValAspPheGluAspProSerSerGluArgLysGluLysGluArgIleProVal 110  
2257 GTAGACTTTTTCAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTG 2316  
111 PheCysIleAspValGluArgAsnAspArgAsnAlaValGlyHisGluProGluHisTyr 130  
2317 TTTTGTATTGATCTTGAAGAAATGATAGAGAGCAGTTCGACAGAGCTGAAATTCG 2376  
131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
2377 TCTGTCTATAGAGATATCTTGAATCTATGTACTTGAATCAAACTAACAAGATTCAT 2436  
151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
2437 GGTGCATTTCTCTGATGCCAGCTTCTCTTAAGAGAGATCATTCGCCCAAAATATGAA 2496  
171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
2497 TTCTTAAAGTCAAGAGGAGGAGTTCGAAGATATCTCAGAAACTTCTGCAGCATCCA 2556  
191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
2557 GAACGTGATGATATGATCAACTCTGTGGCAGATCTTCTTCCCTAATGCTGGGAAACAAA 2616  
211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGly 230  
2617 TTTCTTGTATGATATCTACAGATGTAATCTTGGGAAATATATAAACTGTTCCTGGA 2676  
231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPhelleAsnSer 250  
2677 AAACATATGAAGAGAAAGTGCAGACTTTCGAACCTTTATCATGAATTTCAATTAATCT 2736  
251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
2737 TGTGAGTCTCCAAAGCCATAACCAAGTAGACCAAGTGCACCACTTCAGCCCTCATCTCA 2796  
271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
2797 GAAACCAACACAGAAAGCTTTTCAATGATCTGTTTAAATAATATGCAACCCGTCGTAAT 2856  
291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
2857 ACAGAGAGAAAGCAATATCAGATATTTTATGAGGTGATGACTGTAGAAGGAGTCTAT 2916  
311 AspTyrLeuMetTyrValGlyArgValPheGlnValProAspTyrLeuHisLeu 330  
2917 GATTACCTGATGATAGGACGGGTAGTTTCCAGGTTCCTGACTGCTCATCATCTC 2976  
331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
2977 TTAATGGGAATCGGAATCCTCTTTAAACCAACCCCTGGAAATGTAATGATTACTATCTT 3036  
351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
3037 CAGTGTAACTAGAACACACTATTTTCAGAGCAGCCGTTGGTCTCTCATCAATAACACTCTC 3096  
371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
3097 AGAGATGCTATATTTCTGTGAAACCCCTGAACCTGCTCTCTCCAAAGATAAGCAAAAAGGA 3156  
391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
3157 GCAAAACACACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATT 3216  
411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430

3217 GGTGAAGAAACCCAGTATGAAGCATCAGACTTCTGTTTGAATGCTTACAGCAACCCAGTA 3276  
431 LeuAsnLysGln-----LeuThrTyrValLeuLeuAspIleValIleGlnGluLeu 447  
3277 CTCACCAAGCAGGTGCTTTAGCTGCACTTATGTTTATTTGACATTTGATACAGGAACGTG 3336  
448 PheProGluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
3337 TTTCCAGAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 3390  
RESULT 9  
AK095380 3616 bp mRNA linear PRI 15-JUL-2002  
LOCUS Homo sapiens cDNA FLJ38061 fis, clone CTONG2014966, highly similar  
DEFINITION to SORTING NEXIN 14.  
ACCESSION AK095380  
VERSION AK095380.1 GI:21754626  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
TANIGAMI, A., FUJIWARA, T., SHIBAHARA, T., GOTO, Y., HIRAO, M.,  
SHIMIZU, F., WAKEBE, H., ONO, T., HISHIGAKI, H., WATANABE, T., OZAKI, K.,  
SUGIYAMA, T., IRIE, R., OTSUKI, T., SATO, H., WAKAMATSU, A., ISHII, S.,  
YAMAMOTO, J., ISONO, Y., KAWAI-HIO, Y., SAITO, K., NISHIKAWA, T.,  
KIMURA, K., YAMASHITA, H., MATEO, K., NAKAMURA, Y., SEKINE, M.,  
KIKUCHI, H., KANDA, K., WAGATSUMA, M., MURAKAWA, K., KANEHORI, K.,  
TAKAHASHI-FUJII, A., OSHIMA, A., SUGIYAMA, A., KAWAKAMI, B., SUZUKI, Y.,  
SUGANO, S., NAGAHARI, K., MASUHO, Y., NAGAI, K. and ISOGAI, T.  
NEBO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 3616)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEBO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
FEATURES  
source  
1. 3616  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CTONG2014966"  
/tissue\_type="tongue, tumor tissue"  
/clone\_lib="CTONG2"  
/note="Cloning vector: pME18SFL3"  
BASE COUNT 1078 a 629 c 710 g 1199 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,98e-195 Length: 3616  
Score: 2354.00 Matches: 453  
Percent Similarity: 99.56% Conservative: 1  
Best Local Similarity: 99.34% Mismatches: 1  
Query Match: 96.59% Indels: 1  
DB: 9 Gaps: 0  
US-09-744-313A-1 (1-465) x AK095380 (1-3616)  
QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLys 30  
2080 AGGAACACACAGAAAAGGGGAGATCTTTGGAAATCAGCAATAGTAGCAAAATTAA 2139



QY 31 GlyValPheLysSerThrThrMetGluGluValMetLeuProAsnTyrGlyValAlaGlu 50  
Db 2140 GAGATATTCAAAAGTACCACATGGAGGAGCTATGTTGCTTAATTAATGATGCTGAA 2199  
QY 51 GlyGluAspPheLeuGluGluGluValValMetGluAspSerProValGlu 70  
Db 2200 GGTGAAGATGATTTATTGAAGAGGTATTGTTGTAATGAAGATGATTTCTCCAGTGGAG 2259  
QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTPrLysLeuSerLeuProTyr 90  
Db 2260 GCTGTGAGCACACCTAATCTCCCGAAACCTTCTGCTGATGGAAATTAAGCATTCATAT 2319  
QY 91 ValAspPheGluAspProSerSerGluArgGluGlu-LysLysGluArgLeuProVa 110  
Db 2320 GTAGACTTTTGTGAGTCCCTCTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2379  
QY 110 lPheCysIleAspValGluArgAsnAspArgAlaValGluHisGluProGluHisTr 130  
Db 2380 GTTTTGTATTGATTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2439  
QY 130 pSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheH 150  
Db 2440 GTCTGTCTATAGAAGATCTTGAATCTTATGATCTTGAATCAAACTAACAGAAATTTCA 2499  
QY 150 sGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGl 170  
Db 2500 TGGTGCATTTCTCTGATCCCGAGCTTCTCTTAAGAGGATCATTTGGCCCCAAAATATGA 2559  
QY 170 uPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPr 190  
Db 2560 ATTTCTTAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2619  
QY 190 oGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGl 210  
Db 2620 AGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2679  
QY 210 nPheLeuAspLysLysLeuProAspValAsnLeuGlyLysIleIleLysSerValProGl 230  
Db 2680 ATTTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2739  
QY 230 YLysLeuMetLysGluGluGlyGlnHisLeuGluProPheIleMetAsnPheIleLysLe 250  
Db 2740 AAAAATAATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2799  
QY 250 rCysGluSerProLysProLysProSerArgProGluLeuThrLysLeuSerProThrSe 270  
Db 2800 TTGTGAGTCTCAAGAGCTTAACCAAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2859  
QY 270 rGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAs 290  
Db 2860 AGAAAAACAACAAGAGCTTTTCAATGATCTGTTTAAAAATAATGCAAAACCGTGTGAAAA 2919  
QY 290 nThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTy 310  
Db 2920 TACAG 2979  
QY 310 rAspTyrLeuMetTyrValGlyArgValValPheGluValProAspThrLeuHisLysLe 330  
Db 2980 TGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3039  
QY 330 uLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLe 350  
Db 3040 CTTAATGGGAATCTGATCTCTTTAAAAACACCTGGAAATGATATGATGATGATGATGAT 3099  
QY 350 uGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLe 370  
Db 3100 TCAGTGTAACTAGAACAGCTATTTTCAGGAGCAGCGTTGGTCTCACTCAATAACACTTCT 3159  
QY 370 uArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGl 390  
Db 3160 CAGAGATCTATATTCTGTGAAACACTCAACCTCGCTCTCTCCCAAGATAGCAAAAGG 3219

QY 390 yAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIl 410  
Db 3220 AGCAAAACAGACTTTTGAAGAAATGATGATTAATCATCCAGATCTGTAGTCAAGTGTAT 3279  
QY 410 eGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVa 430  
Db 3280 TGGTGAAGAAACCAAGATGAAGACATCAGACTTCTCTTTGATGGCTTACAGCAACAGT 3339  
QY 430 lLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGl 450  
Db 3340 ACTCAACAGCAGCTGACTTANGTTTATGGACATTTGTATACAGACATCTGTTCCAGA 3399  
QY 450 uLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
Db 3400 GCTCAATAAGGTACAAAGGAAGTTACTCTGTGACATCTTGGATG 3445  
RESULT 10  
LOCUS AF121863 1593 bp mRNA linear PRI 14-SEP-2001  
DEFINITION Homo sapiens sorting nexin 14 (SNX14) mRNA, partial cds.  
ACCESSION AF121863  
VERSION AF121863.1 GI:4689265  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1593)  
AUTHORS Teasdale,R.D., Locci,D., Houghton,F., Karlsson,L. and Gleeson,P.A.  
TITLE A large family of endosome-localized proteins related to sorting nexin 1  
JOURNAL Biochem. J. 358 (Pt 1), 7-16 (2001)  
MEDLINE 21378165  
PUBMED 11485546  
REFERENCE 2 (bases 1 to 1593)  
AUTHORS Teasdale,R.D., Gleeson,P.A. and Karlsson,L.  
TITLE Identification of eleven novel human sorting nexin molecules. A sub-group of the sorting nexin family is associated with the early endosomes  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1593)  
AUTHORS Teasdale,R.D., Gleeson,P.A. and Karlsson,L.  
TITLE Direct Submission  
JOURNAL Submitted (22-JAN-1999) The R.W. Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, CA 92121, USA  
FEATURES  
source  
1..1593  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
gene  
1..1593  
/gene="SNX14"  
CDS  
1..1125  
/gene="SNX14"  
/codon\_start=1  
/product="sorting nexin 14"  
/protein\_id="AAD27836.1"  
/translation="DFPEDPSSEKKEKERIPVFCIDVERNDRAVGHPEHNSVYRR  
VLEFVYLSKLTSEHGAPDQPSKRIIGPKNYEFLKSKEEFQVYLKLOHPPLS  
NSQLADFLSPNGETQFDKILFDVNLGKLIKSPGKLMKEQHLPEFTMFINSC  
ESPKPSPBELTILSPTSNNKKLFNDFKNNANRAENKONKQNTFMFVNTVEGV  
YDLVMYGVRFQVFDWHLHLMGTRILFLMGLTFLYDLYLQLEQLFOHRLVSLI  
TLLDAIPECTEPRSLQDKQKAKQTFENMMYIPDLLVKICIGETKYESIRLLFDG  
LQQPVNLKQLYVLLDIVIQELFPELNNKQKSVTSVSM"

BASE COUNT 533 a 256 c 294 g 510 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.5e-162 Length: 1593  
Score: 1964.00 Matches: 374  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0



[illegible]

383 SerLeuGlnAspLysGlnValAlaLysGlnThrPheGluGluMetMetAsnTyrIle 402  
Db TCCTCCAGATAGCAAAAGGAGCAAAACAGCTTTTGAAGAAATGATGATACATT 320  
403 ProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeu 422  
Db CCAGATCTGTAGTCAAGTCTATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCG 380  
423 PheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIle 442  
Db TTTGATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTTATTGGACATT 440  
443 ValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSerValThr 462  
Db GTGATCAGGAAGTCTTCCAGAGCTCAATPAGGTACAAAGGAGTTACCTCTGTGACA 500  
463 SerTyrMet 465  
Db TCCTGGATG 509

RESULT 12  
BD113855  
LOCUS BD113855 425 bp DNA linear PAT 18-SEP-2002  
DEFINITION EST and encoded human protein.  
ACCESSION BD113855  
VERSION BD113855.1 GI:23208759  
KEYWORDS JP 2002010789-A/5932.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 425)  
AUTHORS Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.  
TITLE EST and encoded human protein  
JOURNAL Patent: JP 2002010789-A 5932 15-JAN-2002;  
GENSET CORP  
COMMENT OS Homo sapiens (human)  
PN JP 2002010789-A/5932  
PD 15-JAN-2002  
PF 07-AUG-2000 JP 2002080989  
PR 05-AUG-1999 US 60/147499  
PI JEAN BAPUIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE  
GIORDANO  
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
C12N15/00  
CC EST and encoded human protein  
FH Key Location/Qualifiers  
FT source 1..425  
FT /organism='Homo sapiens (human)'.  
FEATURES  
source  
1..425  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 134 a 82 c 85 g 124 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4.62e-56 Length: 425  
Score: 743.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 30.49% Indels: 0  
DB: 6 Gaps: 0  
US-09-744-313A-1 (1-465) x BD113855 (1-425)

297 GlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal 316  
Db 1 CAGAAATTATTTATGAGGAGTATGACTGTAGAGGAGTCTATGATTACCTGATGATGTA 60

317 GlyArgValValPheGlnValProAspTyrLeuHisLeuLeuMetGlyThrArgIle 336  
Db GGACGGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAACTCGATC 120  
337 LeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGln 356  
Db CTCCTTAAACAACCCCTGGAAATGTATACTGATTACTATCTTCACTGCTAAACTAGAACAG 180  
357 LeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCys 376  
Db CTATTTCCAGGAGCACCCCTTTGGTCTCACTCATACACTTCTCAGAGATGTATATTCTGT 240  
377 GluAsnThrGluProAspSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396  
Db GAAACACATGACCTCCCTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTGAA 300  
397 GluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr 416  
Db GAAAGCATCAGACTTCTGTTTGTATGAGTCAAGCAACCAAGTACTCTCAACAGCACTGACT 360  
417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436  
Db GAAAGCATCAGACTTCTGTTTGTATGAGTCAAGCAACCAAGTACTCTCAACAGCACTGACT 420  
437 Tyr 437  
Db TAT 423

RESULT 13  
BD026500  
LOCUS BD026500 358 bp DNA linear PAT 27-AUG-2002  
DEFINITION Sequence tag and encoded human protein.  
ACCESSION BD026500  
VERSION BD026500.1 GI:22567723  
KEYWORDS JP 2001269182-A/2746.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 358)  
AUTHORS Edwards,J.B.D.M., Duclair,G. and Jordan,J.Y.  
TITLE Sequence tag and encoded human protein  
JOURNAL Patent: JP 2001269182-A 2746 02-OCT-2001;  
GENSET  
COMMENT OS Homo sapiens (human)  
PN JP 2001269182-A/2746  
PD 02-OCT-2001  
PF 24-FEB-2000 JP 2000118773  
PR 26-FEB-1999 US 60/122487  
PI JEAN BAPUIST DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES  
JORDAN  
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10,  
PC C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
G06F15/40  
CC  
FH Key Location/Qualifiers  
FT CDS 53..358.  
FT Location/Qualifiers  
1..358  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 125 a 63 c 69 g 101 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.2e-46 Length: 358  
Score: 634.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.02% Indels: 0  
DB: 6 Gaps: 0  
US-09-744-313A-1 (1-465) x BD026500 (1-358)

US-09-744-313A-1 (1-465) x BD026500 (1-358)

Qy 216 LeuProaspValAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGlu 235  
 Db 2 CTACCAGATGTAATCTTGGGAAATATATAATCTGTTCTCGAATACTAATGAAGAG 61  
 Qy 236 LysGlyGlnHisLeuGluProPheIleMetAsnPhaIleAsnSerCysGluSerProLys 255  
 Db 62 AAAGGTGAGCATTTGGAAACCTTTTATCATGAATTTTCAATTAATCTTGTGAGTCTCAAG 121  
 Qy 256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnLeuLysLys 275  
 Db 122 CCTAAACCAAGTAGACCAAGACTGACCATTTCTGAGCCCTACTTCAGAAAACACACAGAAG 181  
 Qy 276 LeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
 Db 182 CTTTCATGATCTGTTTAAATAATATGCAACCGTCTGAAATATACAGAGAAAGCAA 241  
 Qy 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315  
 Db 242 AATCAGAAATATTTTATGAGAGTGATGACTGTAGAGAGAGCTCATGATTAACCTGATGTAT 301  
 Qy 316 ValGlyArgValPheGlnValProAspTyrLeuHisLeuMetGlyThr 334  
 Db 302 GTAGCAGCGGTAGTCTTCCAGGTCTCTGACTGCTTCATCATCTCTTAATGGGAACC 358

# RESULT 14

AC116713 278998 bp DNA linear HTG 09-JUN-2003  
 LOCUS Mus musculus clone RP23-118M23, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 7  
 DEFINITION ordered pieces.

AC116713 4 GI:31544102  
 HTG; HTGS PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.

## REFERENCE

AUTHORS Mus musculus (house mouse)  
 TITLE Mus musculus  
 JOURNAL Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 278998)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-118M23  
 2 (bases 1 to 278998)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,  
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fero, S., Ferreira, P., FittHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Govette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kellis, C., Lacroque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McSwan, P., McKernan, K., Meldrim, J., Meneus, L.,  
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norou, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## AUTHORS

Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 278998)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,

Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,  
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hagos, B., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norou, C., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submission@genome.wi.mit.edu](mailto:sequence_submission@genome.wi.mit.edu)

Center project name: L22409

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Center clone name: 118\_M\_23

Qy	172	LeuLysSerLysArgGluGluPheGlnGluTyLeuGlnLysLeuLeuGlnHisProGlu	191
Db	226685	CTCGGAGCCTGAGGAGGAGGTTGTGAGTTTGAGGCCAGCTGAGGCTACATAGCTTT	226626
Qy	192	LeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPhe	211
Db	226625	GCACGCTCTGCATCTCCACGAAAAAACAATAAATAATGAGGGGGCAGAGATGCAACTT	226566
Qy	212	LeuAspLysIleLeuProAspValAsnLeuGly	223
Db	226565	AGTGATAGGATGTTGGCTGGCTTACTTGTAGGCTCTGGATAGTTAGAAATATCAACG	226506
Qy	224	IleIleLysSerValProGlyLysLeuMetLysGluGlyGlnHisLeuGluProPhe	243
Db	226505	ACCATAAGCAGACACCAAAAAAATGGTGAATAA-CAAGAAAGAAAGAGATGCTTT	226447
Qy	244	-----IleMetAsnPheIleAsnSerCys-----	251
Db	226446	ATGCTAGTAATAATTAAGTATGGCAAAAGTGAATACAGCTTTTATTAAATGTTTGA	226387
Qy	252	-----GluSerProLysPro	256
Db	226386	GAGAAATATGGGATTTTTTTTATAGCTTTGGTCATGAATGAGAAAGTTGAAGCA	226327
Qy	257	LysProSerArgProGlu-----LeuThrIleLeuSerProThr	269
Db	226326	GATTTTAACCTTCTAATGTGGCATTTTGTGTATCTTTTAACTAGTATGATCTGTC	226267
Qy	270	SerGluAsnLysLys-----LeuPheAsn	278
Db	226266	AATAATGCAATAAATCTGTCTCTTTATTATTTCACTTATATCTAGTTTCAAC	226207
Qy	279	AspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsn	298
Db	226206	GATCTGTTTAAAGATAATGCAACCGCGCTGAGAACACAGAGGAAACGAAATCAGAAC	226147
Qy	299	TyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal	316
Db	226146	TACTTCAGGAGGTGATGACTGTGGAGGAGTCTATGATTTACCTGATGATGATGTAAG	226087
Qy	316	-----	316
Db	226086	GCTTAATGGTCCAGTTGTTCTTTTGTAGTCTGCTCATGATTCAGATACAGACCTGCTC	226027
Qy	316	-----	316
Db	226026	AACTTTGTGAAGAAATTTCTTTCCCGATTTACTACATTTGGCTTGAACATGACCACTG	225967
Qy	316	-----	316
Db	225966	TTACTTTTATTCATGTATCATCTTCGAGTGACACAGATTTTCAGTTGATTTTAATGGTC	225907
Qy	317	-----GlyArgValValPheGlnValProAspTyrLeuHisLeuLeuMetGlyTh	334
Db	225906	TTCTCTCAGGCGAGTAGTTTTCAGTCCAGACTGCTTCATCATCTTCTATATGGAAC	225847
Qy	334	rArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLe	354
Db	225846	TGGAATCTCTTTAAGAACACCTCGGAAATGTACACAGACTACTTACCTCCAGTCAAGCT	225787
Qy	354	uGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArg-----	371
Db	225786	GGAGCAGCTGTTTCAGGAGCACCGGCTGGTCTCCCTCATCCCTTCTCAGAGTTTCGTA	225727
Qy	371	-----	371
Db	225726	TGACCATCTGTATCATGGTCTACCTTTGTGTTTTTTTTTTTCCAGGTAATTTTAAACC	225667
Qy	371	-----	371
Db	225666	AGCATTTTTTACTAGATATACTTGGTATAAATTTGGAGATAATTTGTATATATCTCC	225607

Qy	372	-----AspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspL	387
Db	225606	TGTACTCATTTTGTAGATGCTATATTTTGTAAATACTGAACCTCGCTCTCTCAAGATA	225547
Qy	387	ysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuV	407
Db	225546	AGCAAAAGGAGCAAAACAGACTTTTGAGAAATGATGAATTACATTTCCAGGTATGATAT	225487
Qy	407	allYs-----CysIle 410	
Db	225486	TTAAATACTCAATTGTATA 225467	
RESULT 15			
AL589666/c			
LOCUS			
DEFINITION	Human DNA sequence from clone RP11-321N4 on chromosome 6, complete		
ACCESSION	AL589666		
VERSION	AL589666.5 GI:14041764		
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 129010)		
AUTHORS	Ramsay, H.		
TITLE	Direct Submision		
JOURNAL	Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
COMMENT	requests: clonerequest@sanger.ac.uk		
	On May 15, 2001 this sequence version replaced gi:13751565.		
	During sequence assembly data is compared from overlapping clones.		
	Where differences are found these are annotated as variations		
	together with a note of the overlapping clone name. Note that the		
	variation annotation may not be found in the sequence submission		
	corresponding to the overlapping clone, as we submit sequences with		
	only a small overlap as described above.		
	This sequence was finished as follows unless otherwise noted: all		
	regions were either double-stranded or sequenced with an alternate		
	chemistry or covered by high quality data (i.e., phred quality >=		
	30); an attempt was made to resolve all sequencing problems, such		
	as compressions and repeats; all regions were covered by at least		
	one plasmid subclone or more than one M13 subclone; and the		
	assembly was confirmed by restriction digest. The following		
	abbreviations are used to associate primary accession numbers given		
	in the feature table with their source databases: Em, EMBL, SW,		
	SWISSPROT, Tr, TrEMBL, Wp, WORMPEP; information on the WORMPEP		
	database can be found at		
	http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence		
	was generated from part of bacterial clone contigs of human		
	chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping		
	Group. Further information can be found at		
	http://www.sanger.ac.uk/HGP/Chr6		
	RP11-321N4 is from the library RPCI-11.2 constructed by the group		
	of Pieter de Jong. For further details see		
	http://www.chori.org/bacpac/home.htm		
	VECTOR: pBAC3.6		
IMPORTANT:	This sequence is not the entire insert of clone		
	RP11-321N4 it may be shorter because we sequence overlapping		
	sections only once, except for a 100 base overlap.		
	The true left end of clone RPI-3U17 is at 128911 in this sequence.		
	The true right end of clone RP11-30P6 is at 100 in this sequence.		
FEATURES			
source			
	1. 129010		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="6"		
	/clone="RP11-321N4"		
	/clone_lib="RPCI-11.2"		
repeat_region			
	2. 192		
	/note="MER21B repeat: matches 3. 207 of consensus"		

repeat\_region 240. .909  
/note="MER21B repeat: matches 30. .715 of consensus"  
repeat\_region 1593. .1682  
/note="MIR repeat: matches 35. .145 of consensus"  
repeat\_region 2030. .2102  
/note="L2 repeat: matches 2681. .2750 of consensus"  
repeat\_region 3459. .3574  
/note="FLAM\_C repeat: matches 1. .117 of consensus"  
repeat\_region 5635. .5670  
/note="LIME repeat: matches 683. .718 of consensus"  
repeat\_region 6887. .6982  
/note="L2 repeat: matches 2616. .2704 of consensus"  
repeat\_region 8284. .8356  
/note="L2 repeat: matches 2639. .2710 of consensus"  
repeat\_region 8380. .8405  
/note="L3 copies 2 mer tt 100% conserved"  
repeat\_region 8440. .8601  
/note="MIR repeat: matches 1. .169 of consensus"  
repeat\_region 8825. .9360  
/note="L2 repeat: matches 1686. .2301 of consensus"  
repeat\_region 9423. .9592  
/note="L2 repeat: matches 2560. .2727 of consensus"  
repeat\_region 12088. .12175  
/note="44 copies 2 mer at 73% conserved"  
repeat\_region 12110. .12177  
/note="L7 copies 4 mer atat 79% conserved"  
repeat\_region 12790. .12905  
/note="L2 repeat: matches 2628. .2750 of consensus"  
repeat\_region 13672. .13923  
/note="L2 repeat: matches 2462. .2725 of consensus"  
repeat\_region 14774. .15053  
/note="AlusX repeat: matches 1. .289 of consensus"  
repeat\_region 15966. .16031  
/note="6 copies 11 mer 77% conserved"  
repeat\_region 17304. .17359  
/note="MIR repeat: matches 196. .251 of consensus"  
repeat\_region 17554. .17682  
/note="MIR repeat: matches 108. .228 of consensus"  
repeat\_region 18366. .18701  
/note="MER2 repeat: matches 1. .345 of consensus"  
repeat\_region 18959. .19248  
/note="AluJo repeat: matches 1. .284 of consensus"  
repeat\_region 20386. .20592  
/note="L2 repeat: matches 70. .288 of consensus"  
repeat\_region 20609. .20778  
/note="LTR28 repeat: matches 848. .1013 of consensus"  
repeat\_region 20747. .21117  
/note="LTR1 repeat: matches 295. .680 of consensus"  
repeat\_region 21163. .21226  
/note="LTR28 repeat: matches 297. .361 of consensus"  
repeat\_region 21187. .21359  
/note="LTR1 repeat: matches 16. .188 of consensus"  
repeat\_region 22191. .22355  
/note="FRAM repeat: matches 4. .163 of consensus"  
repeat\_region 22844. .22991  
/note="MIR repeat: matches 48. .202 of consensus"  
repeat\_region 23155. .23190  
/note="L8 copies 2 mer ga 86% conserved"  
repeat\_region 24390. .24440  
/note="L2 repeat: matches 2657. .2708 of consensus"  
repeat\_region 25082. .25109  
/note="7 copies 4 mer aac 100% conserved"  
repeat\_region 25438. .25744  
/note="L2 repeat: matches 2427. .2744 of consensus"  
repeat\_region 25731. .25917  
/note="MIR repeat: matches 95. .262 of consensus"  
repeat\_region 25964. .26307  
/note="L2 repeat: matches 1986. .2366 of consensus"  
repeat\_region 26526. .26673  
/note="FLAM\_C repeat: matches 1. .142 of consensus"  
repeat\_region 26682. .26737  
/note="L4 copies 4 mer agga 76% conserved"  
repeat\_region 28208. .28394

repeat\_region /note="MIR repeat: matches 12. .196 of consensus"  
29052. .29250  
/note="MIR repeat: matches 48. .251 of consensus"  
30082. .30170  
/note="MIR repeat: matches 50. .145 of consensus"  
30197. .31090  
/note="MER11D repeat: matches 1. .897 of consensus"  
31921. .32067  
/note="FRAM repeat: matches 2. .151 of consensus"  
32545. .32678  
/note="L2 repeat: matches 2578. .2708 of consensus"  
32860. .32895  
/note="9 copies 4 mer ttta 83% conserved"  
32897. .33193  
/note="AlusX repeat: matches 1. .296 of consensus"  
35445. .35478  
/note="MIR repeat: matches 84. .118 of consensus"  
37442. .37598  
/note="MER5B repeat: matches 1. .173 of consensus"  
38659. .38860  
/note="MIR repeat: matches 15. .236 of consensus"  
40552. .40911  
/note="THE1B repeat: matches 1. .364 of consensus"  
40912. .42540  
/note="THE1B-INTERNAL repeat: matches 1. .1580 of consensus"  
42541. .42899  
/note="THE1B repeat: matches 1. .364 of consensus"  
42900. .43231  
/note="MLT1A1 repeat: matches 4. .349 of consensus"  
43407. .43851  
/note="MLT1C repeat: matches 9. .466 of consensus"  
44641. .45020  
/note="W51A repeat: matches 1. .382 of consensus"  
45689. .45988  
/note="AlusG repeat: matches 1. .302 of consensus"  
46220. .46351  
/note="Tiger3b repeat: matches 567. .698 of consensus"  
46358. .46609  
/note="Aluub repeat: matches 40. .290 of consensus"  
46623. .46758  
/note="AluJo/FRAM repeat: matches 170. .301 of consensus"  
46759. .47329  
/note="Tiger3b repeat: matches 2. .540 of consensus"  
47230. .47442  
/note="MER44A repeat: matches 85. .327 of consensus"  
47715. .48005  
/note="L1PA10 repeat: matches 5864. .6153 of consensus"  
48577. .48973  
/note="MLT1B repeat: matches 1. .390 of consensus"  
50633. .50925  
/note="AluY repeat: matches 3. .295 of consensus"  
53828. .55056  
/note="L1MC1 repeat: matches 5023. .6332 of consensus"  
55092. .56081  
/note="L1M4 repeat: matches 3778. .4798 of consensus"  
56074. .56651  
/note="L1MCA repeat: matches 579. .1150 of consensus"  
56652. .56952  
/note="AlusX repeat: matches 1. .301 of consensus"  
56953. .57277  
/note="L1MCA repeat: matches 253. .579 of consensus"  
57280. .57562  
/note="L2 repeat: matches 1590. .1886 of consensus"  
57699. .57766  
/note="L2 repeat: matches 2639. .2710 of consensus"  
59295. .59599  
/note="AlusX repeat: matches 1. .307 of consensus"  
59749. .59811

Alignment Scores:  
Pred. No.:  
Score:

1.24e-38 Length:  
578.00 Matches:  
129010  
129

Percent Similarity: 46.45% Conservative: 2  
 Best Local Similarity: 45.74% Mismatches: 2  
 Query Match: 23.72% Indels: 149  
 DB: 9 Gaps: 2

US-09-744-313A-1 (1-465) x AL589666 (1-129010)

Qy 276 LeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
 Db 58941 CTTTTCATCATCTGTTTAAATAATGCAAAACCGTCTGAAATAACAGAGAGAAAGCAA 58882

Qy 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGluValTyrAspTyrLeuMetTyr 315  
 Db 58881 AATCAGAATATTATTTATGGAGGTGATGCTGTAGAAGAGTCTATGATTACCTGATGTAT 58822

Qy 316 Val----- 316  
 Db 58821 GTAGTAAGATCTAAGTGGTTAAGACTGTTTTTTCATTCATCTTACTTCTCTGTGATG 58762

Qy 316 ----- 316  
 Db 58761 CAAATAATGATCTTGAGTATTGCCCATCATACTGTCITTTTAAATTGAAGATACCTTTT 58702

Qy 316 ----- 316  
 Db 58701 ATAATAATTCATACGATTGCCCTTGATCATATTAAAGTTGCTAGTGTTTACAAATPAAA 58642

Qy 316 ----- 316  
 Db 58641 TATTGACCATGGTCATTGTGTGTTTCTCATCTGGCAAAATTCAAAATACCAATTCATCT 58582

Qy 317 ----- 317  
 Db 58581 GGAACAATGAGGTATCCATTGATCTCTCTCAGACGGGTAGTTTTCAGGTTCTCTGA 58522

Qy 325 pTTPLeuHiHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyr 345  
 Db 58521 CTGGCTTCATCACTCTTTAATGGAACTCGAATCCCTTTTAAACACACCTGGAAATGTA 58462

Qy 345 rThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSe 365  
 Db 58461 TACTGATTACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTC 58402

Qy 365 rLeuIleThrLeuLeuArg----- 371  
 Db 58401 ACTATAACACTTCTCAGAGGTTTGATTTTCTAAAGTGTGTTGTTTCATATTATTGTGTAC 58342

Qy 371 ----- 371  
 Db 58341 TCTATGATTTTAAATTGAATCTAAGAGAGAAATCTTGTGTTTCTGTTTATTAGTA 58282

Qy 371 ----- 371  
 Db 58281 TTTCTACTATGAGAAATCCTTTGTGTATAATAGATGAAATTCAGAAATATTCTTTT 58222

Qy 372 ----- 387  
 Db 58221 TTACTCTTTTATGATCTATTTCTGTGAAACACACGACCTGCTCTCTCCAGATAA 58162

Qy 387 sGlnLysGluValLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuVa 407  
 Db 58161 GCAAAAGGAGCAAAACGACCTTTTGAAGAAATGATGATATACATTCACGCTGTAATATT 58102

Qy 407 Lys 408  
 Db 58101 TAAG 58098

Search completed: January 31, 2004, 04:23:04  
 Job time : 5430 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:52:57 ; Search time 497 Seconds  
(without alignments)

3409.897 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLHFLIPRNTKRGESF.....ELFPELNKVKQEVTSVTSM 465

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113538\_4278/app\_query.fasta\_1.647  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09744313 @CNG 1 1 271 @runat\_30012004\_113538\_4278  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
------------	-------	--------------------	-------	-------------

1	2386	97.9	1716	13	US-10-117-722-922	Sequence 922, App
2	2386	97.9	1716	15	US-10-037-270-922	Sequence 922, App
3	2386	97.9	3470	15	US-10-198-846-13016	Sequence 13016, A
4	2057.5	84.4	1551	13	US-10-117-722-923	Sequence 923, App
5	2057.5	84.4	1551	15	US-10-037-270-923	Sequence 923, App
6	981.5	40.3	725	9	US-09-910-943-712	Sequence 712, App
7	687	28.2	397	11	US-09-918-995-5183	Sequence 5183, App
8	682	28.0	402	10	US-09-983-965-540	Sequence 540, App
9	651	26.7	471	12	US-10-242-535A-25415	Sequence 25415, A
10	544	22.3	508	11	US-09-918-995-20205	Sequence 20205, A
11	505	20.7	305	10	US-09-983-965-355	Sequence 355, App
12	488	20.0	446	12	US-10-242-535A-49594	Sequence 49594, A
13	222	10.8	2720	12	US-10-108-260A-839	Sequence 839, App
14	249.5	10.2	2494	11	US-09-822-846-238	Sequence 238, App
15	248	10.2	280	10	US-09-783-590-4866	Sequence 4866, App
16	235	9.6	2456	10	US-09-784-868-437	Sequence 437, App
17	197	8.1	535	11	US-09-918-995-22530	Sequence 22530, A
18	160.5	6.6	384	10	US-09-783-590-4845	Sequence 4845, App
19	155.5	6.4	2391	12	US-10-293-027-115	Sequence 115, App
20	155.5	6.4	2760	14	US-10-098-841-195	Sequence 195, App
21	152.5	6.3	1294	10	US-09-764-868-198	Sequence 198, App
22	149.5	6.1	2512	9	US-09-784-249-1	Sequence 1, Appl
23	147.5	6.1	1502	9	US-09-822-849A-28	Sequence 28, Appl
24	147.5	6.1	3106	13	US-09-814-353-21324	Sequence 21324, A
25	142	5.8	447	11	US-09-918-995-25777	Sequence 25777, A
26	142	5.8	457	11	US-09-918-995-12082	Sequence 12082, A
27	137.5	5.6	1513	9	US-09-764-868-438	Sequence 438, App
28	137	5.6	1513	9	US-09-925-302-176	Sequence 176, App
29	136.5	5.6	1497	13	US-10-117-722-1021	Sequence 1021, App
30	136.5	5.6	1497	15	US-10-037-270-1021	Sequence 1021, App
31	136.5	5.6	2058	13	US-09-971-392-31	Sequence 31, Appl
32	132	5.4	2004	13	US-09-971-392-30	Sequence 30, Appl
33	128	5.3	1974	13	US-09-873-319-633	Sequence 633, App
34	128	5.3	1974	13	US-09-960-706-971	Sequence 434, App
35	126	5.2	1040	10	US-09-764-868-434	Sequence 316, App
36	126	5.2	1732	10	US-09-864-864-316	Sequence 671, App
37	125	5.1	1718	15	US-10-106-698-671	Sequence 25358, A
38	124	5.1	2628	12	US-10-369-493-25358	Sequence 1, Appl
39	123	5.0	1210	10	US-09-808-701-1	Sequence 1, Appl
40	123	5.0	1210	12	US-10-240-145-27	Sequence 1, Appl
41	123	5.0	1210	15	US-10-233-131-1	Sequence 1, Appl
42	119.5	4.9	1332	12	US-10-369-493-25530	Sequence 25530, A
43	119.5	4.9	3421	13	US-09-814-353-19872	Sequence 19872, A
44	118.5	4.9	3321	13	US-10-032-585-6365	Sequence 6365, App
45	117.5	4.8	418550	12	US-10-292-798-1463	Sequence 1463, App

#### ALIGNMENTS

#### RESULT 1

US-10-117-722-922  
Sequence 922, Application US/10117722  
Publication No. US20030219744A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: NO. US20030219744A1el Nucleic Acids and  
TITLE OF INVENTION: Polyptides  
FILE REFERENCE: 784CIP2BCIP  
CURRENT APPLICATION NUMBER: US/10/117,722  
CURRENT FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 922

290 AenThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
1013 AATAAGAGAGAAACAAAATCAGAAATATTTATGGAGGTGATCACTGTAGAGAGATC 1072  
310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
1073 TATGATTACCTGATGTATGTAGAGGGTAGTTCAGAGTTCCTGACGTGGCTTCATCAT 1132  
330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
1133 CTCCTAATGGGAACTCGAATCCTCTTTAAAAACACCCCTGGAAATGTATACTGATTACTAT 1192  
350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
1193 CTTCACTGTAACTAGAACAGACTATTCAGGAGCACCGTTTGCTCTCACTCAATAACACTT 1252  
370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerIleuGlnAspLysGlnLys 389  
1253 CTCAGAGATGCTATATTCCTGAAACACATGAACCTCGCTCTCTCCAAGATAGCAAAAA 1312  
390 GlyAlaLysGlnThrPheGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
1313 GGAGCAAAAACAGACTTTTGAAGAAATGATGAATTACATTCAGATCTGTGTAGTCAAGTGT 1372  
410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
1373 ATTGGTGAAGAAACCAAGATATGAAGCATCAGACTTCTGTTTGAATGGCTTACAGCAACCA 1432  
430 ValIleuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
1433 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTCATACAGGAACCTGTTTCCA 1492  
450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
1493 GAGCTCAATTAAGGTACAAAAGGAAGTACTCTCTGTGCATCTTGGATG 1540

## RESULT 2

US-10-037-270-922  
: sequence 922 Application US/10037270

; sequence 922, Application 05/19  
: Publication No. US20030104529A1

: GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Chen, Rui-hong

APPLICANT: ZHAO, QING H.  
APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunging

; APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhixiang

APPLICANT: Wang, Zhiwei  
APPLICANT: Tillinghast, J

APPLICANT: DRMANAC, RADOJ

APPLICANT: DUNN, ROBERT  
TITLE OF INVENTION: NO. US

TITLE OF INVENTION: POLYP

FILE REFERENCE: 784CIP2B

1. CURRENT APPLICATION NUMBER

; CURRENT FILING DATE: 2002

; PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE: 2000-04

; PRIOR FILING DATE: 2000-04  
; PRIOR APPLICATION NUMBER:

; PRIOR APPLICATION NUMBER: 2000-01

;; PRIOR FILING DATE: 2000-09-01  
;; NUMBER OF SEQ ID NOS: 1104

; SOFTWARE: pt FL genes Vers

; SEQ ID NO 922

```

; LENGTH: 1716

```

; TYPE: DNA

LENGTH: 1716	1.93e-278	Length:	1716
TYPE: DNA	2386.00	Matches:	456
ORGANISM: Homo sapiens	100.00%	Conservative:	0
FEATURE:	100.00%	Mismatches:	0
NAME/KEY: CDS	97.91%	Indels:	0
LOCATION: (71)..(1543)	13	Gaps:	0
US-10-117-722-922			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-09-744-313A-1 (1-465) x US-10-117-722-922 (1-1716)			
QY 10 PheArgAsnThrGlnLysArgGlyClnSerPheGlyIleSerArgIleGlySerLysValle 29			
DB 173 IIVTGGAAACACACAGAAAGGGAGAAATCATTTGGAAATCAGCAAGATAGGTAGCAAAAT 232			
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49			
DB 233 AAAGAGATATTCAAAGTACCAATGAGGGAGCTATGTTGCCTAATATGTTGGTAGCT 292			
QY 50 GluGlyGluAspAspPheIleGluGlyLeValValMetGluAspAspSerProVal 69			
DB 293 GAAGGTGAAGATGATTTTATTGAAGAGGTATTTGTTGTAATGGAAGATGATTCCTCAGTG 352			
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaIleTolLysLeuSerIlePro 89			
DB 353 GAGGCTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGCTGATGGAATATAGCAATCCA 412			
QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro 109			
DB 413 TATGTAGACTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCT 472			
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValAlGlyHisGluProGluHis 129			
DB 473 GTCTTTTGTATTGATGTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGAACAT 532			
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149			
DB 533 TGTCTCTCTATAGAAGATATCTTGAATTTCTATGTACTTGAATCAAAACTAACAGAAATTT 592			
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169			
DB 593 CATGGTGCAATTTCTGTATGCCAGCTTCCTTCTPAGAGGATCATTTGGCCCCAAAATATAT 652			
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189			
DB 653 GAATCTTAAAGTCAAGAGGGAAGAGTTCACAGAAATATCTACAGAAACTTCTGCAGCAT 712			
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGluThr 209			
DB 713 CCAGAACTGAGTAATAGTCAACTCTTGGCAGACTTTCTTTCCCTTAATGGTGGGGAACA 772			
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229			
DB 773 CAATTTCTTGATAAGTACTACCAAGTGTAAATCTTGGGAAATATTAATAATCTGTCTCT 832			
QY 230 GlyLysLeuMetLysGluLysGlyValGlnHisLeuGluProPheIleMetAsnPheIleAsn 249			
DB 833 GGAAACTAATGAAGAGAAAGGTGAGCATTTGGAACTTTTATCATGATTTTCAATTAAT 892			
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269			
DB 893 TCTTGTGAGTCTCCAAAGCTAAACCAAGTAGACAGAACTGACCACTTCTCACCCCTACT 952			
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289			
DB 953 TCAGAAAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATGAAATGCTGCTGAA 1012			



Dd	1013	AANTCAGAGAGAAAGCAAAAATTCAGAATTATTTTATGGAGGTGATACCTGTAGAGAGTC	1072
Qy	310	TyrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrpLeuHisHis	329
Ddb	1073	TATGATTACCCTGATGTATGTAGACCGGTAGTTTCCAGGTTCTCTGACGTGGGTTTCATCAT	1132
Qy	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
Ddb	1133	CTCTTAATGGGAACCTCGAATCTCTTTAAAAACACCCCTGGAAATGTATACTGATTACTAT	1192
Qy	350	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerIleuIleThrLeu	369
Ddb	1193	CTTCAGTGTAACAATAAGACAGCTATTTTCAGGAGCACCGTTTGGTCTCTCACTATAACACTT	1252
Qy	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
Ddb	1253	CTCAGAGATGCTATATCTCTGTGAACAACACTGAACCTCGCTCTCTCCAAGATAGCAAAA	1312
Qy	390	GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys	409
Ddb	1313	GGAGCAAACAACAGACNTTTGAAGAAATGATGAATTACATTCAGATCTGTTAGTCAAGTG	1372
Qy	410	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro	429
Ddb	1373	ATTGGTAGAAACCAACAGATATGAAGACATCAGACTTCTGTTGTGGCTTACAGCAACCA	1432
Qy	430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449
Ddb	1433	GTACTCAACAAGCAGCTGACTTATGTTTTATTGGACATTGTGTATACAGGAACGTGTTCCA	1492
Qy	450	GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet	465
Ddb	1493	GAGCTCAATAAGGTACAAAGGAAGTTACCTCTGTGCATCTTTGGATG	1540
 RESULT 3 US-10-198-846-13016 ; Sequence 13016, Application US/10198846 ; Publication No. US2003099974A1 ; GENERAL INFORMATION: ; APPLICANT: Lillie, James ; APPLICANT: Xu, Yongyao ; APPLICANT: Wang, Youzhen ; APPLICANT: Steinmann, Kathleen ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS ; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND ; TITLE OF INVENTION: THERAPY OF BREAST CANCER ; FILE REFERENCE: MRI-049 ; CURRENT APPLICATION NUMBER: US/10/198,846 ; CURRENT FILING DATE: 2002-07-18 ; PRIOR APPLICATION NUMBER: 60/306,220 ; PRIOR FILING DATE: 2001-07-18 ; NUMBER OF SEQ ID NOS: 14084 ; SOFTWARE: FastSeq for Windows Version 4.0 ; SEQ ID NO 13016 ; LENGTH: 3470 ; TYPE: DNA ; ORGANISM: Homo sapiens US-10-198-846-13016			
 Alignment Scores:			
Pred. No.:	6,25e-278	Length:	3470
Score:	2386.00	Matches:	456
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.91%	Indels:	0
DB:	15	Gaps:	0
 US-09-744-313A-1 (1-465) x US-10-198-846-13016 (1-3470)			
Qy	10	PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysLeu	29
Dd	1643	TTTCGGNACACACAGAAAAGGGAGATCATTTGGATCAGCAGATAGTAGCAAAATT	1702

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(1543)
US-10-037-270-922

Alignment Scores:
Pred. No.: 1,93e-278 Length: 1716
Score: 2386.00 Matches: 456
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.91% Indels: 0
DB: 15 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-037-270-922 (1-1716)
QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysile 29
DB 173 TTTCGGAAACACACAGAAAGGGAGATCATTTTGGATCAGCAGATAGGTAGCAAAATT 232
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValala 49
DB 233 AAAGGAGTATTCAAAGATACCAATGAGGGAGCTATGTTGCCCTAAATTATGGTGTAGCT 292
QY 50 GluGlyGluAspPheIleGluGluGlyIleValMetGluAspSerProVal 69
DB 293 GAAGGTGAAGATGATTTTATTGAAGAAGGTATTGTTGTAATGGAAGATGATTCCTCAGTG 352
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89
DB 353 GAGGCTGTGAGCACACCTTAATCTCCCCAAACCTTGCTGCGATGGAAATATGACATTCA 412
QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109
DB 413 TATGTAGATCTTTTGGAGATCCCTCTCTGAAGGAGGAGGAAAAGAAAGAAATTCCT 472
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValAlGlyHisGluProGluHis 129
DB 473 GTGTTTGTATTGATGCTTGAAGAAATGATAGAAGAGCAGTGTGGACACGAGCTGAACAT 532
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149
DB 533 TGGTCTGTCTATAGAAGATATCTTGAATTCATGTACTTTGAATCAAAATCAACAGAAATTT 592
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169
DB 593 CATGGTGCAATTCCTGATGCCAGCTTCCTTCTTAAGAGGATCATTTGGCCCCAAAAATAT 652
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189
DB 653 GAATTCCTTAAGTCAAAGAGGGAAGAGTTCAGAAGATATCTACAGAAATCTTCGACGAT 712
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209
DB 713 CCAGAACTGAGTAAATAGTCAACTCTTGGCAGACTTTCTTCCCTAATGTTGGGGAACA 772
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229
DB 773 CAATTCCTTCAATGAATACCTACAGATGTAATCTTGGGAAAAATATAAAATCTGTTCT 832
QY 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249
DB 833 CGAAAACTAATGAAGAGAAAGGTTCAGCATTTTGGAACTCTTTATCATGAAATTTCAATAAT 992
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
DB 893 TCTTGTGAGTCTCCAAAGCCCTAAACCAAGTAGACCACTGACCAATTCAGCCCTACT 952
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
DB 953 TCAGAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAAAATAATGCAACCGTGCTGAA 1012
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309

```



90	TyrValAspPheGluAspProSerSerGluArgIysGluIysIysGluArgIlePro	109
413	TATGTAGACTTTTGTAGGATCCCTCCCTCTCGAAGGAGGAGAAAAAGAAATTCCT	472
110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
473	GTGTTTGTATTTGATGTTTGAAGAATAATATAGAGAGCAGTGTGACACGACCTTGAACAT	532
130	TtpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
533	TGGTCTGCTATAGAGAATATCTTGAAATCTATGTACTTGAATCBAACCTAACAGAAATTT	592
150	HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyr	169
593	CATGGTGCAATTTCTGATGCCAGCTTCTCTTAAGAGGATCAATGGGCCCAAAAAATAT	652
170	GluPheLeuLysSerIysArgGluGluPheGlnGlyTyrLeuGlnLysLeuGlnHis	189
653	GAATCTTAAAGTCAAGAGGAGAGATTCCAGAATAATCTACAGAAACTTCTCTGCAGCAT	712
190	ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
713	CCAGAACTGAGTAATAGTCAACTCTGGCAGACTTCTTCCCTTAATGTGTGGGAACA	772
210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro	229
773	CAATTTCTTGATAGATACTACCAGATGTAATCTTGGGAAATTTATAAAATCTGTTCCT	832
230	GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
833	GGAAAACTAATGAAGAGAGAAAGGTACAGCAATTTGGAACTTTTATCATGAATTTCAATTAAT	892
250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
893	TCTGTGAGTCTCCAAAGCCTAAACCAGTAGACCAAGCTGACCACTCTCGACCCCTACT	952
270	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
953	TCAGAAAAACAAGAAGCTTTTCAATGATCTGTTTAAAAATAATGCAAAACCGTGTGAA	1012
290	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
1013	AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGGAGGTGATGACTCTGACAGGAGTC	1072
310	TyrAspTyrLeuMetTyrValGlyArgValIaPheGlnValProAspTrpLeuHisHis	329
1073	TATGATTACCTGATGTATGTA-----	1093
330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
1093	-----	1093
350	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
1093	-----	1093
370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
1094	-----GATGCTATATTTCTGTGAAAAACATGAACCTCGCTCTCTCCAGATATAGCAAAA	1147
390	GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys	409
1148	GGAGCAAAACAGACTTTTGAAGAATAATGATGATTTACATTCAGATCTGTGTAGTCAAGTGT	1207
410	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro	429
1208	ATTGGTGAAGAAACCAAGTAGAAGCATCAGACTTCTGTGTGATGCTTACAGCAACCA	1267
430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449
1268	GTACTCAACAGCAGCTCACTATGTGTTTATTTGACATTGTGATACAGGCTGTGTTTCCA	1327



Db 545 ATTACTCTGTTAGATGCTATATTTTGTGAAACCGCTGAACACGATCATTACATCTG 604  
Qy 387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
Db 605 AACGACAAAGGCAAGCTTACATTTGAAGAAATGATCGCTATATTCAGATTGATT 664  
Qy 407 ValLysCysIleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeu 426  
Db 665 GGTAAATGATTGCTGATGAAGCTAAATATGAAGCATTCGACTCTCTGTTGGTGATG 724

RESULT 7

US-09-918-995-5183  
; Sequence 5183, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5183  
; LENGTH: 397  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-5183

Alignment Scores:  
Pred. No.: 1,67e-73 Length: 397  
Score: 687.00 Matches: 131  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 28.19% Indels: 0  
DB: 11 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-918-995-5183 (1-397)

Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisLeu 330  
Db 3 GATTACCTGATGATGTAGACGGTAGTTTCCAGGTTCCCTGACTGCTTCATCATCTC 62  
Qy 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
Db 63 TTAATGGAACTCGAATCCCTCTTTAAACACCCCTGGAATGTATCTGATTACTAICTT 122  
Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
Db 123 CAGTGTAACCTAGACAGCTATTTCAGGACACCGTTTGTCTCTCATATAACACTTCTC 182  
Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
Db 183 AGAGTCTATATTTCTGTGAAACACCTGAACTCGCTCTCTCCAGATAAGCAAAAGGA 242  
Qy 391 AlalysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
Db 243 GCAAAACAGACTTTTGAAGAAATGATGAATTACATTCAGATCTGTAGTCAAGTGATT 302  
Qy 411 GlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVal 430  
Db 303 GGTGAGAAACCAAGTATGAAGCATCAGACTCTCTGTTGATGCTTACAGCAACCGATA 362  
Qy 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441  
Db 363 CTCACACAGACGCTGACTTATGTTTATTGGAC 395

RESULT 8

US-09-983-965-540  
; Sequence 540, Application US/09983965  
; Patent No. US20020137160A1

; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 540  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (339)  
; OTHER INFORMATION:  
; OTHER INFORMATION: Clone ID: 64-BOVMS1-017-Q1-E1-H12  
US-09-983-965-540

Alignment Scores:  
Pred. No.: 6.9e-73 Length: 402  
Score: 682.00 Matches: 127  
Percent Similarity: 99.25% Conservative: 5  
Best Local Similarity: 95.49% Mismatches: 1  
Query Match: 27.99% Indels: 0  
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-983-965-540 (1-402)

Qy 205 AsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValaLeuGlyLysIle 224  
Db 3 AATGTTGGGAAACACACAGTTTCTTTGATAAGATACACAGACGTAATCTTGGGAAATT 62  
Qy 225 IleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIle 244  
Db 63 ATAAATCTGTTCTTGAAACTTAATGAAGAGAAAGTCAACATTTGGAACTTTCATC 122  
Qy 245 MetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr 264  
Db 123 ATGATTTTCAATTAATCTTTGTAATCTCCAAAGCTAAACCGAGTAAACAGAACTGACC 182  
Qy 265 IleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsn 284  
Db 183 ATTCTCAGCCCTACCTCAGAGAATAATAAGAAGCTTTTAAATGATCTGTATAAGAAATA 242  
Qy 285 AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMet 304  
Db 243 GCACACCGTCTGAGATAACAGAAAGAGGCAAAATCAGATTAATTTTATGGAATGATG 302  
Qy 305 ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPro 324  
Db 303 ACTGTAGAGAGTCTATGATTACTGATGTATGTATGTATGTATGTATGTATGTATGTAT 362  
Qy 325 AspTrpLeuHisLeuLeuMetGlyThrArgIleLeu 337  
Db 363 GACTGGCTTCATCATCTCTTAATGGAACTCGAATCCTC 401

RESULT 9

US-10-242-535A-25415  
; Sequence 25415, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005

;; CURRENT APPLICATION NUMBER: US/10/242,535A  
;; CURRENT FILING DATE: 2002-09-12  
;; PRIOR APPLICATION NUMBER: US 10/085,783  
;; PRIOR FILING DATE: 2002-02-28  
;; PRIOR APPLICATION NUMBER: US 60/305,340  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/275,017  
;; PRIOR FILING DATE: 2001-03-12  
;; PRIOR APPLICATION NUMBER: US 60/271,955  
;; PRIOR FILING DATE: 2001-02-28  
;; NUMBER OF SEQ ID NOS: 58994  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 25415  
;; LENGTH: 471  
;; TYPE: DNA  
;; ORGANISM: Human  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (17)..(17)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (19)..(19)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (33)..(33)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (49)..(49)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (58)..(59)  
;; OTHER INFORMATION: n is a, c, g, or t

## US-10-242-535A-25415

Alignment Scores:  
Pred. No.: 5,15e-69 Length: 471  
Score: 651.00 Matches: 131  
Percent Similarity: 94.24% Conservative: 0  
Best Local Similarity: 94.24% Mismatches: 7  
Query Match: 26.71% Indels: 1  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-242-535A-25415 (1-471)

QY 328 HisHisLeuLeuMetGlyThrArgIleLeuPhelysAsnThrLeuGluMetTyr-ThrAs 347  
DB 4 CATGACCTGTAAANGNGAACTCGAATCCTTTTAAAAACACCCCTGNAATGTATNNTGA 63  
QY 347 pTyrTyrLeuGlnCysIysLeuGlnLeuPheGlnGluHisArgLeuValSerLeu11 367  
DB 64 TTACTATCTTCAGTGTAAACAGACAGCTATTTTCCAGAGACCCGTTTGTCTCATCTCAT 123  
QY 367 eThrLeuLeuArgaspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLy 387  
DB 124 AACACTTCTCAGATGCTATATCTGTGAAACACCTGACCTCGCTCTCCAGATAA 183  
QY 387 sGlnIysGlyAlaIysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuVal 407  
DB 184 GCMAAAGAGAGCAAAACAGACTTTTGAAGAAATGATGATATACATTCAGATCTGTAGT 243  
QY 407 lIysCysIleGlyGluGluThrIysTyrGluSerIleArgLeuLeuPheAspGlyLeuGl 427  
DB 244 CAAGTGTATTGGTGAAGAAACCAAGTATGAAGATCAGACCTCTGTTTGTGCTTACA 303  
QY 427 nGlnProValLeuAsnIysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLe 447  
DB 304 GCACCACTACTCAACAGACAGCTGACTTATGTTTATGACATTTGTGATCAGAGACT 363  
QY 447 uPheProGluLeuAsnIysValGlnIysGluValThrSerValThrSerTrpMet 465

DB 364 GTTTCAGAGCTCAATAAGGTACAAAGGAAGTTACCTCTGTGACATCTTGGATG 418  
RESULT 10  
US-09-918-995-20205  
;; Sequence 20205, Application US/09918995  
;; Publication No. US20030073623A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc.  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
;; FILE REFERENCE: 20411-756  
;; CURRENT APPLICATION NUMBER: US/09/918,995  
;; CURRENT FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: US/09/235,076  
;; PRIOR FILING DATE: 1999-01-20  
;; NUMBER OF SEQ ID NOS: 38054  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 20205  
;; LENGTH: 508  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-918-995-20205

Alignment Scores:  
Pred. No.: 5,5e-56 Length: 508  
Score: 544.00 Matches: 107  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 22.32% Indels: 0  
DB: 11 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-918-995-20205 (1-508)

QY 359 GlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsn 378  
DB 65 CAGGAGCACCGCTTGGTCTCACTCACTAACAACCTTCTCAGAGATGCTATATTCTGTGAAGAAC 124  
QY 379 ThrGluProArgSerLeuGlnAspIysGlnIysGlyAlaIysGlnThrPheGluGluMet 398  
DB 125 ACTGAACCTCGCTCTCTCAAGATAAGCAAAAGAGCAAAACAGACATTTTGAAGAAATG 184  
QY 399 MetAsnTyrIleProAspLeuValIysCysIleGlyGluGluThrLysTyrGluSer 418  
DB 185 ATGAATTACATTCAGATCTCTTAGTCAAGTGTATTGTTGAAGAAACCAAGTATGAAGC 244  
QY 419 IleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnIysGlnLeuThrTyrVal 438  
DB 245 ATCAGACTTCTGTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTATGTT 304  
QY 439 LeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnIysValGlnIysGluVal 458  
DB 305 TTATTGGACATTTGTGATACAGGAACCTTTTCCAGAGCTCAATAAGGTACAAAGGAAGTT 364  
QY 459 ThrSerValThrSerTrpMet 465  
DB 365 ACCTCTGTGACATCTTGGATG 385

## RESULT 11

US-09-983-965-355  
;; Sequence 355, Application US/09983965  
;; Patent No. US20020137160A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Warren, Wesley C.  
;; APPLICANT: Tao, Nengbing  
;; APPLICANT: Byatt, John C.  
;; APPLICANT: Mathialagan, Nagappan  
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
;; FILE REFERENCE: 37-21(10297)C  
;; CURRENT APPLICATION NUMBER: US/09/983,965  
;; CURRENT FILING DATE: 2001-10-26  
;; PRIOR APPLICATION NUMBER: US 09/465,231

```
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 355
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 42-BOVMS1-019-Q1-E1-C6
US-09-983-965-355

Alignment Scores:
Pred. No.: 1,26e-51 Length: 305
Score: 505.00 Matches: 97
Percent Similarity: 97.03% Conservative: 1
Best Local Similarity: 96.04% Mismatches: 3
Query Match: 20.72% Indels: 0
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-983-965-355 (1-305)
QY 180 GlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuAla 199
DB 1 CAGGAATATCTGCAGAACTCTGCACATCCAGATTGATGATCAGTCACTTCTGGCA 60

QY 200 AspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysLeuProAspVal 219
DB 61 GATTTCTTCTCCCAATGTTGGGAAACACAGTTTCTGATTAAGTACTTACCAGACGTA 120

QY 220 AsnLeuGlyLysLleLleLysSerValProGlyLysLeuMetLysGlyGlnHis 239
DB 121 AATCTTGGGAAAATATATAATCTGTTCTGGAAAACATAATGAAGAGAAAGTCAACAT 180

QY 240 LeuGluProPheLleMetAsnPheLleAsnSerCysGluSerProLysProSer 259
DB 181 TTGAACCTTCATCATGAATTTCTTAATCTTGGAATCTCAAGCCTTAACCGAGT 240

QY 260 ArgProGluLeuThrLleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAsp 279
DB 241 AAACAGACTGACCAATCTCAGGCCTACATCAGAGATATATAACAGCTTTTATGAT 300

QY 280 Leu 280
DB 301 CTG 303

RESULT 12
US-10-242-535A-49594
; Sequence 49594, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 49594
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-49594

; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 355
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 42-BOVMS1-019-Q1-E1-C6
US-09-983-965-355

Alignment Scores:
Pred. No.: 2.73e-49 Length: 446
Score: 488.00 Matches: 95
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.02% Indels: 0
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-242-535A-49594 (1-446)
QY 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390
DB 1 AGAGATGCTATATCTCTGTAACACACTGAACCTCGCTCTCTCCAAGATAAGCAAAAGGA 60

QY 391 AlaLysGlnThrPheGluGluMetMetAsnTyrLleProAspLeuLeuValLysCysLle 410
DB 61 GCAAAACACAGACTTTTGAAGAAATGATGAATTACATTCAGATCTGTGTAGTCACTGTATT 120

QY 411 GlyGluGluThrLysTyrGluSerLleArgLeuLeuPheAspGlyLeuGlnProVal 430
DB 121 GGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTGTGATGCTTACAGCAACAGTA 180

QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspLleValleGlnGluLeuPheProGlu 450
DB 181 CTCACACAGCAGCTGACTTATGTTTATTTGACATTTGTGATACAGGAACCTGTTCCAGAG 240

QY 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465
DB 241 CTCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 285

RESULT 13
US-10-108-260A-839
; Sequence 839, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 839
; LENGTH: 2720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-839

Alignment Scores:
Pred. No.: 1.41e-20 Length: 2720
Score: 262.00 Matches: 98
Percent Similarity: 42.65% Conservative: 79
Best Local Similarity: 23.61% Mismatches: 154
Query Match: 10.75% Indels: 84
DB: 12 Gaps: 15

US-09-744-313A-1 (1-465) x US-10-108-260A-839 (1-2720)
QY 56 lleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal 72
DB 1133 TTGAAGGATGAATAATCTTAATAGAGAAAGACGACAGACCTTCAGCTGCACATGGCA 1192

QY 73 SerThrProAsnThrProArgAsnLeuAlaAlaTrpLysLleSerIleProTyr-ValAsp 92
DB 1193 AGAACGGATTGGTGGTGTGAAAACCTTGGCATGTGGAAGAGCTCATCACCAGTGGAGAG 1252

QY 93 PhePheGluAspProSerSerGluArgLysGlyLysGluArgLleProValPheCys 112
DB 1253 GTTACAGAA-----GAGATGGTGAGCAATTCCTGATGTTACTTT 1291

QY 113 lleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerVal 132
DB 1292 GTCATGGTAAAGCCACAAAGAA-----GTTGGAGGAGTTGAAGACTAGAACTGGACGGTC 1345
```

QY 133 TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAla 152  
Db 1346 CCAGAGGCTCAGCAGATTTCAGATTTCACCGGAACTCAGTGAGTGCCTCTCT 1405  
QY 153 PheProAspAlaGlnLeuProSerLysArgilelleGlyProLysAsnTyrGlu----- 170  
Db 1406 TTAATAAAGTCAGTTCCTCTCTTAGCAAGCTGCTTTCATAATCTATAGATCAAAAG 1465  
QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 1466 TTTATGAAAGTCGAGATCAATTAATAATTTTACAGATCTGCTTTTCAGATGAA 1525  
QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
Db 1526 AGACTGTGTGAGAGTGAAGCTTTTATGCTTTCAGGCTTCTCTGACTACCTCAAG 1585  
QY 211 PheLeuAspLysileLeuProAspValAsnLeuGlyLysilelleLysSerValProGly 230  
Db 1586 GTTATCGAC-----GTGAGGGG 1603  
QY 231 LysLeuMetLysGlnLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
Db 1604 -----AAAAAATCTTTTTCATTATCTCTGTTTGT----- 1636  
QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 1636 ----- 1636  
QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290  
Db 1637 -----GAAGACTTCTCGGACTTCTTC-----TCCACAGAGGAGGAG 1678  
QY 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 1679 ACAGAGGAGGACAGTGCCTGTCAGATTATGTTGAT-----GATGTGGATGGAGAA 1732  
QY 311 AspTyrLeu-----MetTyrValGlyArgValValPheGlnValPro---- 324  
Db 1733 GACGCTTGGCTGAACCATGTTTCATGTTGATGTTGGGAG--ATTITGAATCTCGAGGA 1789  
QY 325 -----AspTyrLeuHisLeuMetGly---ThrArgIleLeuPheLysAsnThr 341  
Db 1790 ATGTTTAAATGGGAGAGACATTAATGCTCTGCTCAGTGCATCTTTGGAGNACC 1849  
QY 342 LeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHis 361  
Db 1850 ATC-----AACAAACAAATCCGGACACAGTGCAGTGGATTTCAGTGAGCAA 1897  
QY 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePhe-----Cys 376  
Db 1898 ATGTTGGTTTACTACATCAATTAATTTCCGGATGCTTTTGGCCAAATGGGAAGTTGGCA 1957  
QY 377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396  
Db 1958 CCACGACCAATCAGAAAGCAAGCAAGTGCAGAAACAAACAGAGACAGACAGCAA 2017  
QY 397 GluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr 416  
Db 2018 AAGCTGCTTGAACAACTTCCAGATATGCTTCAGAGCCTTGTGGACAGCAAAATGCCCGC 2077  
QY 417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436  
Db 2078 CACGGTATAATAAATAATTCATGCACTGCAAGAAACAGAGCCAAACAGCATCTCTTA 2137  
QY 437 TyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeu 451  
Db 2138 TATGCGCTGATGGAAGTCTGCTAATTAATGAATGTGCTGAGCTG 2182

RESULT 14

US-09-822-846-238

; Sequence 238, Application US/09822846

; Publication No. US20030027139A1

GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Bowman, Michael R.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakar  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6400  
; CURRENT APPLICATION NUMBER: US/09/822,846  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,605  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 629  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 238  
; LENGTH: 2494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-846-238

Alignment Scores:  
Pred. No.: 4e-19 Length: 2494  
Score: 249.50 Matches: 100  
Percent Similarity: 42.75% Conservative: 77  
Best Local Similarity: 24.15% Mismatches: 155  
Query Match: 10.24% Indels: 83  
DB: 11 Gaps: 14

US-09-744-313A-1 (1-465) x US-09-822-846-238 (1-2494)

QY 56 lleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal 72  
Db 908 TTGAAGATGAATAATCTTAATAGAGAAAGACGACAGACCTTCAGTCCATGGCA 967  
QY 73 SerThrProAsnThrProArgAsnLeuAlaAlaIleLysIleSerIleProTyrValAsp 92  
Db 968 AGAACGGATTGGTGTGTGAAACCTTGGCATGTGGAAAGCCTCCATCCAGTGGAGAG 1027  
QY 93 PhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCys 112  
Db 1028 GTTACAGAA-----GAGAATGGTGAGCAATTGCCATGTACTTT 1066  
QY 113 lleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisIleTyrSerVal 132  
Db 1067 GTCTGGTAAGCCTTACAAGAA-----GTTGGAGAGATTGAACCTAAGAACTGACGGTC 1120  
QY 133 TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAla 152  
Db 1121 CCCAGAGGCTCAGCAGATTTCAGAAATTTACCGGAAACTCAGTGAAGTGCCTCTCT 1180  
QY 153 PheProAspAlaGlnLeuProSerLysArgIlelleGlyProLysAsnTyrGlu----- 170  
Db 1181 TTAATAAAGTCCAGTTCCTCTCTTAGCAAGTGCCTTTCAAATCTATAGATCAAAAG 1240  
QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 1241 TTTATGAAAGTCCGAGAAATCAATTAATAAGTTTTTACAGAAATCTCTCTTCAGATGAA 1300  
QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210



Db 1301 AGACTGTGTCAGAGTGAAGCACTTTATGCTTCTGAGCCCTTCTCTGACTACTCTCAAG 1360  
Qy 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValProGly 230  
Db 1361 GTTATGAC-----GTGAGGG 1378  
Qy 231 LysLeuMetLysGlyGlnHisLeuGluProPheLeuMetAsnPheLeuAsnSer 250  
Db 1379 -----AAAAAATCTTTTTCATTATCTCTCTTTG----- 1411  
Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 1411 ----- 1411  
Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
Db 1412 -----GAAAGACTTCTCCGACTTCTTC-----TCCACACGAGGAGGAG 1453  
Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 1454 ACAGAGGAGGAGCAGTGCCTGTCAGATTATGTTGAT-----GATGTGATGGAGGAAA 1507  
Qy 311 AspTyrLeu-----MetTyrValGlyArgValPheGlnValPro----- 324  
Db 1508 GAGCGCTTGCTGAACCATGCTTTTCATGTTGATGGGAG-----ATTTTGAACCTCGAGGA 1564  
Qy 325 -----AspTyrLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu 342  
Db 1565 ATGTTTAAATGGGTGAGAGAACATTAAATGTC-----CTCGTTCAGTCACTTTG 1615  
Qy 343 GluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnPheGlnGluHisArg 362  
Db 1616 GAGAG-ACCATCAACAACAATCCGGGACACAGTCAGTGGATTTTCAGTGAGCAATG 1674  
Qy 363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePhe-----CysGlu 377  
Db 1675 TTGTTTACTACATATTTTCCGGATGCTTTTGGCCAAATCGGAAGTTGGCACCA 1734  
Qy 378 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGlu 397  
Db 1735 CCGACCAACAATCAGAGCAAGAGCAAGTTCAGGAACAACAACAGAGACACAGCAAAAG 1794  
Qy 398 MetMetAsnTyrIleProAspLeuValLysCysIleGlyGluThrLysTyrGlu 417  
Db 1795 CTGCTTGAAACATTCAGATATGCTTCAGAGCCTTGTGGACAGCAAAATGCCGCCAC 1854  
Qy 418 SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyr 437  
Db 1855 GGTATATAAATAATTTCAATGCACTGCAAGAACAGAGCCCAAGCATCTGTTATAT 1914  
Qy 438 ValLeuLeuAspIleValIleGlnGluLeuPheProGluLeu 451  
Db 1915 GCGCTGATGAAGTCTGCTGAATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1956

RESULT 15  
US-09-783-590-4866  
; Sequence 4866, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; PRIOR FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21

; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4866  
; LENGTH: 280  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (17)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (50)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (92)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (127)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (146)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (171)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (208)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (220)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (233)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (253)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (274)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (275)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-4866

Alignment Scores:  
Pred. No.: 1,58e-20 Length: 280  
Score: 248.00 Matches: 63  
Percent Similarity: 77.91% Conservative: 4  
Best Local Similarity: 73.26% Mismatches: 12  
Query Match: 10.18% Indels: 7  
DB: Gaps: 2

US-09-744-313A-1 (1-465) x US-09-783-590-4866 (1-280)

Qy 264 ThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsn-AspLeuPheLysAs 283  
Db 3 ACCATTCTCAGCCCNACTTCAGAAAACAACAAGAGCTTTCAATGANTCTGTTTAAAAA 62  
Qy 283 AsnAla-AsnArgAlaGluAsnThrGluArg--LysGlnAsnGlnAsnTyrPheMetG1 302  
Db 63 TAATGCCAAACCGCTGAAATAACAGAGNGATAAGACAAATCAGAAATATTATTATGGA 122  
Qy 302 uValMetThrValGluGlyValTyrAspTyr---LeuMetTyrValGlyArgValValPh 321  
Db 123 GGTGNTGACTAGAGGAGGTGNTGAATACCTGAATGTGATGTAGNCGGAGTATTT 182  
Qy 321 eGlnValProAspTyrLeuHisLeuLeuMetGlyThrArgIleLeu----PheLysA 340  
Db 193 CCAGNTCTGCTGCTGCTTCTCATCTATCTTAATGGGNGCTACGTATCCNCTTTTAAAA 242

Oy 340 snThrieu 342  
| | | | |  
Db 243 ACACCTG 250

Search completed: January 31, 2004, 06:31:34  
Job time : 523 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 15:56:14 ; Search time 125 Seconds  
(without alignments)  
7033.882 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1-gtataaaactccaaaagt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO\_NUC  
Gapop 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/prodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/PCTUS-COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	76.9	1716	4	US-09-620-312D-922
2	918	46.1	1551	4	US-09-620-312D-923
3	20	1.0	271	2	US-08-731-272A-29
4	20	1.0	7015	3	US-09-177-249-6
5	19	1.0	731	1	US-08-451-405A-2
6	19	1.0	2060	1	US-07-721-761A-31
7	19	1.0	2060	1	US-07-978-687-31
8	19	1.0	2060	5	PCT-US91-01750-2
9	19	1.0	2060	5	PCT-US91-05801-31
10	19	1.0	2081	5	PCT-US91-01750-3
11	19	1.0	3440	1	US-08-471-791-27
12	19	1.0	3440	5	PCT-US91-01746-27
13	19	1.0	3826	4	US-09-220-132-51
14	18	0.9	141	3	US-08-737-078A-1
15	18	0.9	141	5	PCT-US94-04705-1
16	18	0.9	243	4	US-09-134-001C-671
17	18	0.9	544	4	US-09-280-116-247
18	18	0.9	632	4	US-09-489-847-56
19	18	0.9	1001	4	US-09-841-638-339
20	18	0.9	1001	4	US-09-641-638-363
21	18	0.9	1001	4	US-09-641-638-600
22	18	0.9	1395	4	US-09-134-001C-651
23	18	0.9	1926	4	US-09-613-303-50
24	18	0.9	2356	1	US-08-105-483-222
25	18	0.9	2356	1	US-08-220-151-75
26	18	0.9	2356	1	US-08-413-118-75
27	18	0.9	2356	1	US-08-224-657-51

C 28	18	0.9	2356	1	US-08-709-209-222	Sequence 222, App
C 29	18	0.9	2356	1	US-08-458-101-222	Sequence 222, App
C 30	18	0.9	2356	2	US-08-184-009-78	Sequence 78, Appl
C 31	18	0.9	2356	2	US-08-417-210A-68	Sequence 68, Appl
C 32	18	0.9	2356	2	US-08-458-356-78	Sequence 78, Appl
C 33	18	0.9	2356	3	US-08-473-446-75	Sequence 75, Appl
C 34	18	0.9	2356	3	US-08-460-736-78	Sequence 78, Appl
C 35	18	0.9	2356	4	US-09-354-138-51	Sequence 51, Appl
C 36	18	0.9	2356	4	US-09-535-370-78	Sequence 78, Appl
C 37	18	0.9	2760	1	US-08-101-593-1	Sequence 1, Appl
C 38	18	0.9	2760	1	US-08-101-593-3	Sequence 3, Appl
C 39	18	0.9	3475	4	US-08-924-629C-15	Sequence 15, Appl
C 40	18	0.9	4072	3	US-09-272-496-7	Sequence 7, Appl
C 41	18	0.9	4206	4	US-09-302-620B-81	Sequence 81, Appl
C 42	18	0.9	4206	4	US-09-912-161-3	Sequence 3, Appl
C 43	18	0.9	5385	4	US-08-961-527-77	Sequence 77, Appl
C 44	18	0.9	5796	4	US-09-366-715-4	Sequence 4, Appl
C 45	18	0.9	10614	1	US-08-135-511-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1  
US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 8569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1543)  
US-09-620-312D-922

Query Match 76.9%; Score 1532; DB 4; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 147 GGACACACAGAAAGGGGAGATCATTTGGATCAGACAGATAGGTAGCAATTAAG 206  
Db 177 GGACACACAGAAAGGGGAGATCATTTGGATCAGACAGATAGGTAGCAATTAAG 236

207 GAGTATTTCAAAAGTACCAATGAGGAGGCTATGTTGCTTAATTTATGTTGTAGCTGAAG 266  
237 GAGTATTTCAAAAGTACCAATGAGGAGGCTATGTTGCTTAATTTATGTTGTAGCTGAAG 296  
267 GTGAAGATGATTTTATTTGAAGAGGTATTTGTTGAATGGAAGATGATTTCTCCAGTGGAGG 326  
297 GTGAAGATGATTTTATTTGAAGAGGTATTTGTTGAATGGAAGATGATTTCTCCAGTGGAGG 356  
327 CTGTGAGCAACCTAATATCTCCCGAAAACCTTGCTGCTGATGGAATAATTTAGCATTCATATG 386  
357 CTGTGAGCAACCTAATATCTCCCGAAAACCTTGCTGCTGATGGAATAATTTAGCATTCATATG 416  
387 TAGACTTTTTTGGAGTCCCTCTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 446  
417 TAGACTTTTTTGGAGTCCCTCTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 476  
447 TTTGATTTGATGTTGAAGAAATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 506  
477 TTTGATTTGATGTTGAAGAAATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 536  
507 CTGCTATGAGAGATATCTTGAAATCTATGATCTGAATCAAAATCAAAATCAAAATCAAAAT 566  
537 CTGCTATGAGAGATATCTTGAAATCTATGATCTGAATCAAAATCAAAATCAAAATCAAAAT 596  
567 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAT 626  
597 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAT 656  
627 TCTTAAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 686  
657 TCTTAAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 716  
687 AACTGATGATAGTCAACTTTCTGGCAGACTTTCTTTCCCTTAATGGTGGGAGGAGGAGGAGG 746  
717 AACTGATGATAGTCAACTTTCTGGCAGACTTTCTTTCCCTTAATGGTGGGAGGAGGAGGAGG 776  
747 TTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 806  
777 TTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836  
807 AACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 866  
837 AACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 896  
867 GTGAGTCTCAAAAGCTTAAACCAAGTAGACCAAGTACCACTTCTCAGCCCTACTTTCAG 926  
897 GTGAGTCTCAAAAGCTTAAACCAAGTAGACCAAGTACCACTTCTCAGCCCTACTTTCAG 956  
927 AAAACCAAGAGAGCTTTTCAATGATCTGTTTAAAAAATATGCAAAACCGTCTGAAATA 986  
957 AAAACCAAGAGAGCTTTTCAATGATCTGTTTAAAAAATATGCAAAACCGTCTGAAATA 1016  
987 CAG 1046  
1017 CAG 1076  
1047 ATTACCTGATGATGATGAGAGAGGAGTATTTTCCAGGTTCTGACTGGTTCATCTCTCT 1106  
1077 ATTACCTGATGATGATGAGAGAGGAGTATTTTCCAGGTTCTGACTGGTTCATCTCTCT 1136  
1107 TAATGGGAACCTGAATCTTTTAAAAACACCTCGCTCTCTCCAAAGATAGCAAAAGAGGAG 1166  
1137 TAATGGGAACCTGAATCTTTTAAAAACACCTCGCTCTCTCCAAAGATAGCAAAAGAGGAG 1196  
1167 AGTGTAACTAGAACAGCTATTTTCCAGGAGCACCCTGTTTGGTCTCAGTCACTAATCACTCTCA 1226  
1197 AGTGTAACTAGAACAGCTATTTTCCAGGAGCACCCTGTTTGGTCTCAGTCACTAATCACTCTCA 1256  
1227 GAGATGCTATATCTGTGAG 1286  
1257 GAGATGCTATATCTGTGAG 1316  
1287 CAAACACAGACTTTTGAAG 1346

1317 CAAACACAGACTTTTGAAG 1376  
1347 GTGAAG 1406  
1377 GTGAAG 1436  
1407 TCAACAG 1466  
1437 TCAACAG 1496  
1467 TCAACAG 1526  
1497 TCAACAG 1556  
1527 GTATAG 1586  
1557 GTATAG 1616  
1587 GTATATTTCTTAT 1646  
1617 GTATATTTCTTAT 1676  
1647 AATAAG 1678  
1677 AATAAG 1708

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP28  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1378)  
US-09-620-312D-923

Query Match 46.1%; Score 918; DB 4; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

147 GGAACACACAGAAAGGGAGGAATCATTTGGATCAGCAGATAGGTAGCAAAATTAAG 206  
 177 GGAACACACAGAAAGGGAGGAATCATTTGGATCAGCAGATAGGTAGCAAAATTAAG 236  
 207 GAGTATTCAAAAGTACCAAAAGGAGGAGCTATGCTGCTAATATATGCTGAGCTGAAG 266  
 237 GAGTATTCAAAAGTACCAAAAGGAGGAGCTATGCTGCTAATATATGCTGAGCTGAAG 296  
 267 GTGAAGATGATTTATTTGAAGAGGATTTGTTGTAATGAAGATGATTTCTCCAGTGGAGG 326  
 297 GTGAAGATGATTTATTTGAAGAGGATTTGTTGTAATGAAGATGATTTCTCCAGTGGAGG 356  
 327 CTGTGAGCACACCTAATATCTCCCGAAACCTTCTGCTGATGGAATATGACATTCATATG 386  
 357 CTGTGAGCACACCTAATATCTCCCGAAACCTTCTGCTGATGGAATATGACATTCATATG 416  
 387 TAGACTTTTTCAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 446  
 417 TAGACTTTTTCAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 476  
 447 TTGTATTCATGTTGAAAGAAATGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 506  
 477 TTGTATTCATGTTGAAAGAAATGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 536  
 507 CTGTCTATAGAAGATATCTTGAATCTTGTATGATCAAAATCAAAATCAAAATCAAAAT 566  
 537 CTGTCTATAGAAGATATCTTGAATCTTGTATGATCAAAATCAAAATCAAAATCAAAAT 596  
 567 GTGCATTTCTGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 626  
 597 GTGCATTTCTGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 656  
 627 TCTTAAGTCAAGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 686  
 657 TCTTAAGTCAAGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 716  
 687 AACTGAGTATAGTCAACTTTCTGCGAGAGCTTTCTTCCCTTAATGTTGGGAGGAGGAGGAG 746  
 717 AACTGAGTATAGTCAACTTTCTGCGAGAGCTTTCTTCCCTTAATGTTGGGAGGAGGAGGAG 776  
 747 TTCTTGATAGATACATACAGAGTAAATCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 806  
 777 TTCTTGATAGATACATACAGAGTAAATCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 836  
 807 AACTATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 866  
 837 AACTATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 896  
 867 GTGAGTCTCCAAAGCCTTAAACCAAGTAGACCAAGTCACTTCTCAGCCCTACTTCAG 926  
 897 GTGAGTCTCCAAAGCCTTAAACCAAGTAGACCAAGTCACTTCTCAGCCCTACTTCAG 956  
 927 AAAACAAACAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGGAATA 986  
 957 AAAACAAACAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGGAATA 1016  
 987 CAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1046  
 1017 CAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1076  
 1047 ATTACCTGATGATGAG 1064  
 1077 ATTACCTGATGATGAG 1094

RESULT 3  
 US-08-731-272A-29  
 ; Sequence 29, Application US/08731272A  
 ; Patent No. 5837463  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tanaka, Torahiko  
 ; APPLICANT: Katoh, No. 5837463uyuki

APPLICANT: Shimotohno, Kunitada  
 TITLE OF INVENTION: NUCLEIC ACID OF C TYPE HEPATITIS VIRUS  
 TITLE OF INVENTION: DERIVATION AND PROCESS FOR DETECTION VIRUS USING SAID  
 TITLE OF INVENTION: NUCLEIC ACID  
 NUMBER OF SEQUENCES: 29  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SUGRUE, MION, ZINN, MACPEAK & SEAS  
 STREET: 2100 Pennsylvania Avenue, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/731,272A  
 FILING DATE: 11-OCT-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP HEI. 7/268700  
 FILING DATE: 17-OCT-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Nakamura, Dean H.  
 REGISTRATION NUMBER: 33,981  
 REFERENCE/DOCKET NUMBER: Q-43086  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)293-7060  
 TELEFAX: (202)293-7860  
 INFORMATION FOR SEQ ID NO: 29:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 271 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: RNA (genomic)  
 US-08-731-272A-29  
 Query Match 1.08; Score 20; DB 2; Length 271;  
 Best Local Similarity 25.08; Pred. No. 16;  
 Matches 5; Conservative 15; Mismatches 0; Indels 0; Gaps 0;  
 Qy 54 AACTGTTTCTCTTTTCTTTTCTTTT 73  
 Db 74 AACGUGUUUUUUUUUUUUUUUU 93  
 RESULT 4  
 US-09-177-249-6  
 ; Sequence 6, Application US/09177249  
 ; Patent No. 6229064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer, Robert L.  
 ; APPLICANT: Chad, Nir  
 ; APPLICANT: Kiyosue, Tomohiro  
 ; APPLICANT: Yadegari, Ramin  
 ; APPLICANT: Margossian, Linda  
 ; APPLICANT: Harada, John  
 ; APPLICANT: Goldberg, Robert B.  
 ; APPLICANT: The Regents of the University of California  
 ; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
 ; FILE REFERENCE: 023070-0861200S  
 ; CURRENT APPLICATION NUMBER: US/09/177,249  
 ; CURRENT FILING DATE: 1998-10-22  
 ; EARLIER APPLICATION NUMBER: US 09/071,838  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 324  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 6  
 ; LENGTH: 7015

TYPE: DNA  
ORGANISM: Arabidopsis sp.  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(7014)  
OTHER INFORMATION: fertilization-independent endospore 1 (FIE1)  
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)..(7015)  
OTHER INFORMATION: fertilization-independent endospore 1 (FIE1)  
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3)..(7013)  
OTHER INFORMATION: fertilization-independent endospore 1 (FIE1)  
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3  
US-09-177-249-6

Query Match 1.0%; Score 20; DB 3; Length 7015;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 TTTTCTTTTAAATGTTTT 84  
|||||  
DB 5523 TTTTCTTTTAAATGTTTT 5542

RESULT 5  
US-08-451-405A-2  
Sequence 2, Application US/08451405A  
Patent No. 5736358  
GENERAL INFORMATION:  
APPLICANT: FASEL, NICOLAS JOSEPH  
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE  
TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND  
TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE WEBB LAW FIRM  
STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE  
CITY: PITTSBURGH  
STATE: PENNSYLVANIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 15219-1818  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISK  
COMPUTER: Midwest Micro 486-50  
OPERATING SYSTEM: DOS  
SOFTWARE: WORDPERFECT 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,405A  
FILING DATE: 26-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,273  
FILING DATE: 15-JAN-1993  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: UNKNOWN  
US-08-451-405A-2

Query Match 1.0%; Score 19; DB 1; Length 731;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1907 TTTTCTTTTAAATGTTTT 1925  
|||||  
DB 71 TTTTCTTTTAAATGTTTT 89

RESULT 6  
US-07-721-761A-31  
Sequence 31, Application US/07721761A  
Patent No. 5475099  
GENERAL INFORMATION:  
APPLICANT: Vic. C. Knauf  
APPLICANT: Gregory A. Thompson  
TITLE OF INVENTION: Plant Fatty Acid Synthases  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/721,761A  
FILING DATE: 19910626  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/568,493  
FILING DATE: 15-AUGUST-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Laesen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
REFERENCE/DOCKET NUMBER: CGNE 76-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-07-721-761A-31

Query Match 1.0%; Score 19; DB 1; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989  
|||||  
DB 113 TTTTAAATAATAAAAAA 131

RESULT 7  
US-07-978-687-31  
Sequence 31, Application US/07978687  
Patent No. 5510255  
GENERAL INFORMATION:  
APPLICANT: Vic. C. Knauf  
APPLICANT: Gregory A. Thompson  
TITLE OF INVENTION: Plant Fatty Acid Synthases  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
FILING DATE: FEBRUARY 1, 1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05801  
FILING DATE: 15-AUGUST-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/568,493  
FILING DATE: 15-AUGUST-1990  
PRIOR APPLICATION DATA: 07/721,761  
FILING DATE: 26-JUNE-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
REFERENCE/DOCKET NUMBER: CGNE 76-2 WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-07-978-687-31

Query Match 1.0%; Score 19; DB 1; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1971 TTTTAAATAATAAAAAA 1989  
Db 113 TTTTAAATAATAAAAAA 131

RESULT 8  
PCT-US91-01750-2  
Sequence 2, Application PC/TUS9101750  
GENERAL INFORMATION:  
APPLICANT: KNAUF, VIC C.  
APPLICANT: KRIDL, JEAN C.  
APPLICANT: SCHERER, DONNA E.  
TITLE OF INVENTION: Novel Sequences Preferentially  
TITLE OF INVENTION: Expressed in Early Seed  
TITLE OF INVENTION: Development and Methods  
TITLE OF INVENTION: Related Thereto  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01750  
FILING DATE: 19910314  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/494,722  
FILING DATE: 16-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
REFERENCE/DOCKET NUMBER: CGNE 68WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 916-753-6313  
TELEFAX: 916-753-1510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
PCT-US91-01750-2

Query Match 1.0%; Score 19; DB 5; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1971 TTTTAAATAATAAAAAA 1989  
Db 113 TTTTAAATAATAAAAAA 131

RESULT 9  
PCT-US91-05801-31  
Sequence 31, Application PC/TUS9105801  
GENERAL INFORMATION:  
APPLICANT: VIC. C. Knauf  
APPLICANT: Gregory A. Thompson  
TITLE OF INVENTION: Plant Patty Acid Synthases  
NUMBER OF SEQUENCES:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05801  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/568,493  
FILING DATE: 15-AUGUST-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/721,761  
FILING DATE: 26-JUNE-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
REFERENCE/DOCKET NUMBER: CGNE 76-2 WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double

```
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
PCT-US91-05801-31

Query Match      1.0%; Score 19; DB 5; Length 2060;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 113 TTTTAAATAATAAAAAA 131

RESULT 10
PCT-US91-01750-3
; Sequence 3, Application PC/TUS9101750
; GENERAL INFORMATION:
; APPLICANT: KNAUF, VIC C.
; APPLICANT: KRIDL, JEAN C.
; TITLE OF INVENTION: Novel Sequences Preferentially
; TITLE OF INVENTION: Expressed In Early Seed
; TITLE OF INVENTION: Development and Methods
; TITLE OF INVENTION: Related Thereto
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 5.0
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01750
; FILING DATE: 19910314
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/494,722
; FILING DATE: 16-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 68WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 916-753-6313
; TELEFAX: 916-753-1510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2081 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
PCT-US91-01750-3

Query Match      1.0%; Score 19; DB 5; Length 2081;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 113 TTTTAAATAATAAAAAA 131

RESULT 11
US-08-471-791-27

; Sequence 27, Application US/08471791
; Patent No. 5723595
; GENERAL INFORMATION:
; APPLICANT: Thompson, Gregory A
; APPLICANT: Knauf, Vic C
; TITLE OF INVENTION: Plant Desaturases-Compositions
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: California
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1 (a)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,791
; FILING DATE: 6-JUNE-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/762,762
; FILING DATE: 16-SEPT-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01746
; FILING DATE: 14-MAR-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/615,784
; FILING DATE: 14-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/567,373
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/494,106
; FILING DATE: 16-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Lassen, Elizabeth
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 69-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; TELEX: 350370 CGNE
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-471-791-27

Query Match      1.0%; Score 19; DB 1; Length 3440;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 344 TTTTAAATAATAAAAAA 362

RESULT 12
PCT-US91-01746-27
```



; Sequence 27, Application PC/TUS9101746  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Gregory A  
; APPLICANT: Knauf, Vic C  
; TITLE OF INVENTION: Plant Desaturases-Compositions and Uses  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calgene, Inc.  
; STREET: 1920 Fifth Street  
; CITY: Davis  
; STATE: California  
; COUNTRY: USA  
; ZIP: 95616  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 6.0.7  
; SOFTWARE: Microsoft Word 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/01746  
; FILING DATE: 19910314  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/615,784  
; FILING DATE: 14-NOV-1990  
; APPLICATION NUMBER: 07/567,373  
; FILING DATE: 13-AUG-1990  
; APPLICATION NUMBER: 07/494,106  
; FILING DATE: 16-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lassen, Elizabeth  
; REGISTRATION NUMBER: 31,845  
; NAME: Donna E. Scherer  
; REGISTRATION NUMBER: 34,719  
; REFERENCE/DOCKET NUMBER: CGNE 69-3 WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (916) 753-6313  
; TELEFAX: (916) 753-1510  
; TELEX: 350370 CGNE  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3440 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; PCT-US91-01746-27

Query Match 1.0%; Score 19; DB 5; Length 3440;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989  
DB 344 TTTTAAATAATAAAAAA 362

RESULT 13  
US-09-220-132-51  
; Sequence 51, Application US/09220132  
; Patent No. 6506607  
; GENERAL INFORMATION:  
; APPLICANT: Sylvian, Andrew W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
; FILE REFERENCE: 07334-074001  
; CURRENT APPLICATION NUMBER: US/09/220,132  
; CURRENT FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: US 60/079,303  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: US 60/068,821  
; PRIOR FILING DATE: 1997-12-24  
; NUMBER OF SEQ ID NOS: 191

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 51  
; LENGTH: 3826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-220-132-51

Query Match 1.0%; Score 19; DB 4; Length 3826;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1974 TTAATAATAAAAAA 1992  
DB 3793 TTAATAATAAAAAA 3811

RESULT 14  
US-08-737-078A-1/c  
; Sequence 1, Application US/08737078A  
; Patent No. 6027934  
; GENERAL INFORMATION:  
; APPLICANT: Powell, Curtis  
; TITLE OF INVENTION: VACCINE FOR, DIAGNOSTIC ASSAY FOR AND  
; METHOD OF TREATING PARASITIC HEMOFLAGELLATE PROTOZOA  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe  
; STREET: 660 White Plains Road  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10591-5144  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: System 7.5  
; SOFTWARE: Wordperfect 3.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/737,078A  
; FILING DATE: 16-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/04931  
; FILING DATE: 11-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/714,464  
; FILING DATE: 13-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/365,413  
; FILING DATE: 13-JUN-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kurt G. Briscoe  
; REGISTRATION NUMBER: 33,141  
; REFERENCE/DOCKET NUMBER: POWELL 203-PCT-KGB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914) 332-1700  
; TELEFAX: (914) 332-1844  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-737-078A-1

Query Match 0.9%; Score 18; DB 3; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1975 TAATAATAAAAAA 1992  
DB 67 TAATAATAAAAAA 50

RESULT 15  
 PCT-US94-04706-1/c  
 ; Sequence 1, Application PC/TUS9404706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Powell, Curtis  
 ; TITLE OF INVENTION: VACCINE FOR DIAGNOSTIC ASSAY FOR AND  
 ; TITLE OF INVENTION: METHOD OF TREATING PARASITIC  
 ; TITLE OF INVENTION: HEMOFLAGELLATE PROTOZOA  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Spring Horn Kramer & Woods  
 ; STREET: 660 White Plains Road  
 ; CITY: Tarrytown  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10591-5144  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
 ; COMPUTER: Apple Macintosh  
 ; OPERATING SYSTEM: System 7.0  
 ; SOFTWARE: WordPerfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/04706  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/04931  
 ; FILING DATE: 11-JUN-92  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/714,464  
 ; FILING DATE: 13-JUN-91  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/365,413  
 ; FILING DATE: 13-JUN-89  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kurt G. Briscoe  
 ; REGISTRATION NUMBER: 33,141  
 ; REFERENCE/DOCKET NUMBER: POWELL 203-PCT-KGB  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (914) 332-1700  
 ; TELEFAX: (914) 332-1844  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 141 nucleotides  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 PCT-US94-04706-1

Query Match 0.9%; Score 18; DB 5; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1975 TAATAATAAAAAAAAAA 1992  
 Db 67 TAATAATAAAAAAAAAA 50

Search completed: January 31, 2004, 19:34:48  
 Job time : 128 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 31, 2004, 13:58:54 ; Search time 7413 Seconds  
(without alignments)  
10993.117 Million cell updates/sec  
Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1 gtagaaactcaaaagt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 2888711 seqs, 20454813386 residues  
Word size: 0

Total number of hits satisfying chosen parameters: 5777422  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.cm.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.in.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1992	100.0	1992	6	AX054819	Sequence
2	1593	80.0	1593	9	AF121863	AX054819 Homo sapi
3	1538	77.2	3038	9	BC005110	BC005110 Homo sapi
4	1369	68.7	2661	9	AY044865	AY044865 Homo sapi
5	1354	68.0	2925	9	BC046520	BC046520 Homo sapi
6	1315	66.0	3615	9	AK095380	AK095380 Homo sapi
7	1296	65.1	3145	6	AX512835	AX512835 Sequence
8	1296	65.1	3145	9	AK000362	AK000362 Homo sapi
9	1171	58.8	3576	9	AK026479	AK026479 Homo sapi
10	816	41.0	968	9	HS420561	HS420561 Homo sapi
11	512	25.7	129010	9	AL589666	AL589666 Human DNA
12	425	21.3	425	6	BD113855	BD113855 EST and e
13	385	19.3	451	6	AX331384	AX331384 Sequence
14	357	17.9	358	6	BD026500	BD026500 Sequence
15	143	7.2	454	11	G30543	G30543 human STS S
16	81	4.1	1782	10	BC043328	BC043328 Mus muscu
17	78	3.9	580	6	AX387326	AX387326 Sequence
18	76	3.8	311	6	AX185131	AX185131 Sequence
19	76	3.8	313	6	AX186466	AX186466 Sequence
20	76	3.8	27898	2	AC116713	AC116713 Mus muscu
21	74	3.7	299	6	AX188027	AX188027 Sequence
22	74	3.7	224551	2	AC111832	AC111832 Rattus no
23	74	3.7	278375	2	AC130093	AC130093 Rattus no
24	35	1.8	160411	2	AC135935	AC135935 Rattus no
25	25	1.3	181833	2	AC141405	AC141405 Homo sapi
26	25	1.3	194034	9	AP005901	AP005901 Homo sapi
27	25	1.3	211492	2	AC140877	AC140877 Homo sapi
28	25	1.3	216811	2	AC140808	AC140808 Homo sapi
29	24	1.2	92628	9	AL355472	AL355472 Human DNA
30	24	1.2	145239	9	AC138701	AC138701 Homo sapi
31	24	1.2	149779	2	AL929118	AL929118 Homo sapi
32	24	1.2	150936	2	BX255899	BX255899 Danio rer
33	24	1.2	152668	5	AL935302	AL935302 Zebrafish
34	24	1.2	162533	2	BX537304	BX537304 Danio rer
35	24	1.2	168498	2	AL938847	AL938847 Homo sapi
36	24	1.2	169546	2	AC004157	AC004157 Plasmodi
37	24	1.2	174018	2	BX004989	BX004989 Danio rer
38	24	1.2	177789	2	BX470207	BX470207 Danio rer
39	24	1.2	179140	2	AC140860	AC140860 Homo sapi
40	24	1.2	179143	2	BX511195	BX511195 Danio rer
41	24	1.2	179577	5	AL928843	AL928843 Zebrafish
42	24	1.2	180715	2	BX005052	BX005052 Danio rer
43	24	1.2	217607	10	AL691509	AL691509 Mouse DNA
44	24	1.2	217698	2	BX248129	BX248129 Danio rer
45	24	1.2	224934	2	BX511028	BX511028 Danio rer

ALIGNMENTS

RESULT 1  
AX054819  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AX054819  
Sequence 3 from Patent WO0073334.  
AX054819  
AX054819.1 GI:12228268  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Yue, H., Tang, Y. T. and Azimzai, Y.  
Human sorting nexins  
Patent: WO 0073334-A 3 07-DEC-2000;

AX054819  
1992 bp  
DNA  
linear  
PAT 13-JAN-2001

Incyte Genomics, Inc. (US)  
Location/Qualifiers  
1. .1992  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 671 a 308 c 372 g 641 t  
ORIGIN

Query Match 100.0%; Score 1992; DB 6; Length 1992;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATGAAACTCAAAAGTTATGTTAGCTATACCTTTAGTTTATCATTTTCAAACTGTT 60  
DB 1 GTATGAAACTCAAAAGTTATGTTAGCTATACCTTTAGTTTATCATTTTCAAACTGTT 60

QY 61 TTTCTTTTAAATGCTTTTCAATAAAGTACAGCTGTTGTGATAAATTTGCTATGTA 120  
DB 61 TTTCTTTTAAATGCTTTTCAATAAAGTACAGCTGTTGTGATAAATTTGCTATGTA 120

QY 121 CTTGATACATTTTGTGTTTAAATATTCAGGAACACACAGAAAGGGGAGAAATCATTTTGGAA 180  
DB 121 CTTGATACATTTTGTGTTTAAATATTCAGGAACACACAGAAAGGGGAGAAATCATTTTGGAA 180

QY 181 CAGCAGAAATAGTAGCAAAATTAAGAGATTTCAAAAGTACCAATGAGAGGAGCTAT 240  
DB 181 CAGCAGAAATAGTAGCAAAATTAAGAGATTTCAAAAGTACCAATGAGAGGAGCTAT 240

QY 241 GTTGCTTAATTTAGTGTAGCTGAGGTGAAGATGATTTTATTTGAAGAGGATTTGTTGT 300  
DB 241 GTTGCTTAATTTAGTGTAGCTGAGGTGAAGATGATTTTATTTGAAGAGGATTTGTTGT 300

QY 301 AATGGAAGATGATTTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCGGAAACCTTGC 360  
DB 301 AATGGAAGATGATTTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCGGAAACCTTGC 360

QY 361 TGCATGAAATATAGCAATCCATATGTAGACTTTTGTGAGATCCCTCTCTGAAAGGAA 420  
DB 361 TGCATGAAATATAGCAATCCATATGTAGACTTTTGTGAGATCCCTCTCTGAAAGGAA 420

QY 421 GGAGAAAGAAAGAAATTCCTGTTTGTATTTGATTTGTAAGAAATGATAGAGAGC 480  
DB 421 GGAGAAAGAAAGAAATTCCTGTTTGTATTTGATTTGTAAGAAATGATAGAGAGC 480

QY 481 AGTTGGACAGAGCTCAACATTTGCTGTATAGAAATATCTTGAATCTCTATGACT 540  
DB 481 AGTTGGACAGAGCTCAACATTTGCTGTATAGAAATATCTTGAATCTCTATGACT 540

QY 541 TGAATCAAAACTAACAGAAATTCATGTTGCAATTCCTGATGCCAGCTTCCTTCTAAGAG 600  
DB 541 TGAATCAAAACTAACAGAAATTCATGTTGCAATTCCTGATGCCAGCTTCCTTCTAAGAG 600

QY 601 GATCATTTGGCCCAAAATTAAGAAATTTAAAGTCAAGAGGAGAGGTTCCAGAAATA 660  
DB 601 GATCATTTGGCCCAAAATTAAGAAATTTAAAGTCAAGAGGAGAGGTTCCAGAAATA 660

QY 661 TCTACAGAACTTCTGAGCATCCAGAACTGAGTAAGTACCTCTGCGACATTTCT 720  
DB 661 TCTACAGAACTTCTGAGCATCCAGAACTGAGTAAGTACCTCTGCGACATTTCT 720

QY 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGATAGATACTACAGATGTAATCTTGG 780  
DB 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGATAGATACTACAGATGTAATCTTGG 780

QY 781 GAAATATTAATATCTGTTCTGGAATACTAATGAAGAGAAAGGTCAGATTTGGAAC 840  
DB 781 GAAATATTAATATCTGTTCTGGAATACTAATGAAGAGAAAGGTCAGATTTGGAAC 840

QY 841 TTTTATCATGATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTTAACCAAGTAGACCA 900  
DB 841 TTTTATCATGATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTTAACCAAGTAGACCA 900

QY 901 ACTGACCATTCTCAGCCTACTTTCAAGAAACAAAGAGCTTTTCAATGATCTGTTAA 960  
DB 901 ACTGACCATTCTCAGCCTACTTTCAAGAAACAAAGAGCTTTTCAATGATCTGTTAA 960

QY 961 AAATAATGCAAAACCGTCTGAAATAACAGAGAGAAAGCAAAATCAAGATTTATTTATGGA 1020  
DB 961 AAATAATGCAAAACCGTCTGAAATAACAGAGAGAAAGCAAAATCAAGATTTATTTATGGA 1020

QY 1021 GGTGATGACTGTAGAGGAGTCTATGATTTACCTGATGTATGTAGGACGGTAGTTTCCA 1080  
DB 1021 GGTGATGACTGTAGAGGAGTCTATGATTTACCTGATGTATGTAGGACGGTAGTTTCCA 1080

QY 1081 GGTTCCTGACTGGCTTCATCATCTCTTAATGGAACCTCGAATCTCTTTAAAAACACCT 1140  
DB 1081 GGTTCCTGACTGGCTTCATCATCTCTTAATGGAACCTCGAATCTCTTTAAAAACACCT 1140

QY 1141 GGAATGTATACCTGATTTACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGAGACCG 1200  
DB 1141 GGAATGTATACCTGATTTACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGAGACCG 1200

QY 1201 TTTGGTCTCATCTONAAACATCTTCAGAGATGCTATATTTCTGTGAAAAACCTGAACTCG 1260  
DB 1201 TTTGGTCTCATCTONAAACATCTTCAGAGATGCTATATTTCTGTGAAAAACCTGAACTCG 1260

QY 1261 CTCTCTCCAGATAAGCAAAAGAGCAAAAGAGCAAAAGAGCACTTTTGAAGAAATGATTAAT 1320  
DB 1261 CTCTCTCCAGATAAGCAAAAGAGCAAAAGAGCAAAAGAGCACTTTTGAAGAAATGATTAAT 1320

QY 1321 TCCAGATCTGTTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATGAAGCAATCAGATTCT 1380  
DB 1321 TCCAGATCTGTTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATGAAGCAATCAGATTCT 1380

QY 1381 GTTTGATGGCTTACAGCAACAGTACTCAACAGAGAGCTGACTTATGTTTATTTGGACAT 1440  
DB 1381 GTTTGATGGCTTACAGCAACAGTACTCAACAGAGAGCTGACTTATGTTTATTTGGACAT 1440

QY 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAGAGTACCTCTGTGAC 1500  
DB 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAGAGTACCTCTGTGAC 1500

QY 1501 ATCTTGATGTAAACACTTGGATTTGGTATAGATAACCAATTTGAAATTTCTGCTGTGCG 1560  
DB 1501 ATCTTGATGTAAACACTTGGATTTGGTATAGATAACCAATTTGAAATTTCTGCTGTGCG 1560

QY 1561 AGGTGGTGTAGAAATTTACTTTTTGGGTATATTTCTTATATATATATATGATCATGCTGTC 1620  
DB 1561 AGGTGGTGTAGAAATTTACTTTTTGGGTATATTTCTTATATATATATATGATCATGCTGTC 1620

QY 1621 TGAATTTTGTAGTATTTTGTGTTTAAAGAGCTAAACCAAACTTTAATGATTTAAAGT 1680  
DB 1621 TGAATTTTGTAGTATTTTGTGTTTAAAGAGCTAAACCAAACTTTAATGATTTAAAGT 1680

QY 1681 GATTTGATCTCATAGTCTTTTCAATTTGCTAGTGTATCCAAATTTTATAGAACATTAAGT 1740  
DB 1681 GATTTGATCTCATAGTCTTTTCAATTTGCTAGTGTATCCAAATTTTATAGAACATTAAGT 1740

QY 1741 CACTTTGTTATTTGCCATTTTAAAAAGAGAAATTCATTAATGATGTTATGCAAAACAGATAA 1800  
DB 1741 CACTTTGTTATTTGCCATTTTAAAAAGAGAAATTCATTAATGATGTTATGCAAAACAGATAA 1800

QY 1801 GACTGATAAACTTCGTATTTGTATAGCTTTGAAATAAATATGCTGTATGAGAGAAACAG 1860  
DB 1801 GACTGATAAACTTCGTATTTGTATAGCTTTGAAATAAATATGCTGTATGAGAGAAACAG 1860

QY 1861 GAAATAAGATCTGATTTTCTTAGAGTTAATATTTTATAGTATGTTTCTTTTTTTTTT 1920  
DB 1861 GAAATAAGATCTGATTTTCTTAGAGTTAATATTTTATAGTATGTTTCTTTTTTTTTT 1920

QY 1921 ATTTTGTACATAGTTTAACTGTGATCTATAAATAAGCACTCTATATGAGTTTTTATAA 1980  
DB 1921 ATTTTGTACATAGTTTAACTGTGATCTATAAATAAGCACTCTATATGAGTTTTTATAA 1980

QY 1981 TAAAAAATAAATAA 1992



```
Db 1261 TAAAGACTAACAACTTAATGATTAAAGTGATTGAGTCTCATAGTCTTTCAATTGCT 1320
QY 1709 AGCTGATCCAAATTTATTAGACATAAGTCACTGTTATTGCAATTTTAAAGAGA 1768
Db 1321 AGCTGATCCAAATTTATTAGACATAAGTCACTGTTATTGCAATTTTAAAGAGA 1380
QY 1769 AAATTCATATGATGTTATGGCAAAACAGATAAGACTGATAAACTTCGTATTGATAGCTT 1828
Db 1381 AAATTCATATGATGTTATGGCAAAACAGATAAGACTGATAAACTTCGTATTGATAGCTT 1440
QY 1829 TGAATAATATGCTAGTATGAGAGAAACAGATAAGACTGATGATGTTTCTTAGAGTTAA 1888
Db 1441 TGAATAATATGCTAGTATGAGAGAAACAGATAAGACTGATGATGTTTCTTAGAGTTAA 1500
QY 1889 TATATTTTAGTAGATGCTGTTTCTTTTCTTTTATTTTGTACACTGATGATGATCTA 1948
Db 1501 TATATTTTAGTAGATGCTGTTTCTTTTCTTTTATTTTGTACACTGATGATGATCTA 1560
QY 1949 TAAATAAGCATCTATATGAGTTTCTTTTCTTTTATTTTGTACACTGATGATGATCTA 1981
Db 1561 TAAATAAGCATCTATATGAGTTTCTTTTCTTTTATTTTGTACACTGATGATGATCTA 1593

RESULT 3
LOCUS BC005110 3038 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:13217 IMAGE:3959086, mRNA, complete cds.
ACCESSION BC005110
VERSION BC005110.1 GI:13477272
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3038)
Strausberg, R.
Direct Submission
Submitted (26-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Kzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Nataasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 18 Row: m Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
Location/Qualifiers
1..3038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:13217 IMAGE:3959086"
/class_type="Placenta, Chorioncarcinoma"
/clone_lib="NIH_MGC_21"
```

CDS

```
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
189..2849
/codon_start=1
/product="Unknown (protein for MGC:13217)"
/protein_id="AAH05110.1"
/db_xref="GI:13477273"
/translation="MGQKQKRLDLVDGRIICROYPLFCFLLCISAAALLNRYIHI
LMTFWSFVAVVTFYCSLGSLLPNIFFTIKYKPKQLGLQELFPOGSCAVCGKVKC
KEHPSLLENVOPWLDLKITSSKVASLSEVDIPSIITKKLLKAMKHIEVIVKAROK
VMTPELQAAALEYGPPELHVALSRDELHKLRLTELLFVLPKATDCESLTL
IRILSGSVLPFLSDFLADPTVNHLLIIFDPSPEKATEPASELVLPLOKFAERN
KPSVLKELKQIREQDLDFRPMFLKQEGAVHVLQCLTVEBFDVNLKLMRCLFAY
MLSLHEELQIKYKTYCLDRESIDKRPDIIVEIQRIAEGPIYDFVVKLQTMRLPSFE
EHVLSLENVFTPMFCHSDYFQRLRGAESPTNSKSKNVAENTQKRGESGIGRISGIS
KGVKSTMEGAMLPNVYAGEDEDFIERSIVMEDDSPEAVSTENTPRNLAAAKIS
IPVDFEPPSPSERKKEKRIIPVFCIDVERNDRRAVGEHEPMSVVRVLEFVLESK
LTFHGAFPDPAQLPSKLIIGPKNYIEFKSKREGEFQYELQKLOHPELSOLLALFSL
PNGGTQFLDKLIPDVNLGKIISKVPEKLMKRGQHLPEFIMNFINSCESPKPKSRP
ELTILSPSTENKKLFDNLFKNNAEAENTKONQNYFMEVNTVEGVVDYMDLIPCE
VPOVPDWLHLLMGTRILPKNTLEMTVDYVLOCKLQELFOEHLVSLIYLLADALPCE
NTEPSRLQDKQKAKQTFEMMNYIPDLVKICIGETKYESIRLLFDGLQQPVNLQKL
TYVLLDIVIQELFPFLNKVKQEVTSVTSWM"
```

BASE COUNT  
ORIGIN

```
Query Match 77.2%; Score 1538; DB 9; Length 3038;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 CAGGAACACACAGAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATTA 204
Db 1481 CAGGAACACACAGAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATTA 1540
QY 205 AGGAGTATTCAAAGTACCAATGAGGAGGAGCTATGTTGCCTAAATATGCTGAGCTGA 264
Db 1541 AGGAGTATTCAAAGTACCAATGAGGAGGAGCTATGTTGCCTAAATATGCTGAGCTGA 1600
QY 265 AGGTGAGATGATTTTATGGAAGAGTATGTTGTAATCGAAGATGATTTCTCCAGTGA 324
Db 1601 AGGTGAGATGATTTTATGGAAGAGTATGTTGTAATCGAAGATGATTTCTCCAGTGA 1660
QY 325 GCCTGTGAGCACACTAATACTCCCCGAAACCTTCTGTCATGGAATAATAGCATTCATA 384
Db 1661 GCCTGTGAGCACACTAATACTCCCCGAAACCTTCTGTCATGGAATAATAGCATTCATA 1720
QY 385 TGTAGACTTTTTCAGGATCCCTCTCTGAAAGAGGAGAGAAAAGAAAGAAATTCCTGT 444
Db 1721 TGTAGACTTTTTCAGGATCCCTCTCTGAAAGAGGAGAGAAAAGAAAGAAATTCCTGT 1780
QY 445 GTTTTGTATTGATGTTGAAAGAAATGATAGAGAGGAGTTGGACAGAGCTGAAACATTG 504
Db 1781 GTTTTGTATTGATGTTGAAAGAAATGATAGAGAGGAGTTGGACAGAGCTGAAACATTG 1840
QY 505 GTCTGCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATTTCA 564
Db 1841 GTCTGCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATTTCA 1900
QY 565 TGGTGCAATTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATATGA 624
Db 1901 TGGTGCAATTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATATGA 1960
QY 625 ATTCTTAAGTCAAGAGGAGAGAGTTCCAGAAATATCTACAGAAATCTTCGAGCATCC 684
Db 1961 ATTCTTAAGTCAAGAGGAGAGAGTTCCAGAAATATCTACAGAAATCTTCGAGCATCC 2020
QY 685 AGAAGTGAATAGTCACTTCTGGCAGACTTCTTTCCCTTAATGTTGGGGGAAACACA 744
Db 2021 AGAAGTGAATAGTCACTTCTGGCAGACTTCTTTCCCTTAATGTTGGGGGAAACACA 2080
QY 745 ATTTCTTGATAAGATACCTACCAGATGTAATCTTTGGGAAATATATAAAATCTGTTCTGG 804
Db 2081 ATTTCTTGATAAGATACCTACCAGATGTAATCTTTGGGAAATATATAAAATCTGTTCTGG 2140
```







db	1740		GTCTGTCCTATAGAAGATATCTTGAATTCCTATGTCTTGAATCAAAATCTAAACGAATTTCA	1799
Qy	565		TGTTGCATTTCTTGATGCCAGCTTCCTCTTAAGAGGATCATTGGCCCCAAAAATTTATGA	624
Db	1800		TGTTGCATTTCTTGATGCCAGCTTCCTCTTAAGAGGATCATTGGCCCCAAAAATTTATGA	1859
Qy	625		ATTCCTTAAAGTCAAAAGAGGAAGTGTCCAAAGATATCTACAGAAATTCCTGCGAGCATCC	684
Db	1860		ATTCCTTAAAGTCAAAAGAGGAAGTGTCCAAAGATATCTACAGAAATTCCTGCGAGCATCC	1919
Qy	685		AGAACTCAGTAATAGTCAACTTCTGCGCAGACTTTCTTCCCTCTAAATGGTGGGGAAACACA	744
Db	1920		AGAACTCAGTAATAGTCAACTTCTGCGCAGACTTTCTTCCCTCTAAATGGTGGGGAAACACA	1979
Qy	745		ATTTCTTTGATAAGATATCTACAGATGTAAATCTTTGGGAAATTTATAAAATCTGTTCCTGG	804
Db	1980		ATTTCTTTGATAAGATATCTACAGATGTAAATCTTTGGGAAATTTATAAAATCTGTTCCTGG	2039
Qy	805		AAAACTAATGAAGAGAAAGGTACAGATTTGGAAACCTTTTATCAGAAATTTCAATTAATTC	864
Db	2040		AAAACTAATGAAGAGAAAGGTACAGATTTGGAAACCTTTTATCAGAAATTTCAATTAATTC	2099
Qy	865		TTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGACTGACCAATTCCTCAGCCCTACTTC	924
Db	2100		TTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGACTGACCAATTCCTCAGCCCTACTTC	2159
Qy	925		AGAAAACAAACAAGAGCTTTTCAATGATCTGTCTTAAAAATAATGCAAAACCGTGTCTGAAAA	984
Db	2160		AGAAAACAAACAAGAGCTTTTCAATGATCTGTCTTAAAAATAATGCAAAACCGTGTCTGAAAA	2219
Qy	985		TACAGAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAAGGAGTCTA	1044
Db	2220		TACAGAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAAGGAGTCTA	2279
Qy	1045		TGATTACTGATGTATGTAGACGGGTAGTTTTCCAGGTTTCCTGACCTGGCTTCATCATCT	1104
Db	2280		TGATTACTGATGTATGTAGACGGGTAGTTTTCCAGGTTTCCTGACCTGGCTTCATCATCT	2339
Qy	1105		CTTAAATGGGAACCTCGAATCCCTTTTAAAAACAACCTCGGAATGTATCTGATTACTATCT	1164
Db	2340		CTTAAATGGGAACCTCGAATCCCTTTTAAAAACAACCTCGGAATGTATCTGATTACTATCT	2399
Qy	1165		TCAGTGTAAACTAGAACAGCTATTTCAGGAGCACCGTTTGGTCTCTCACTATAACACTTCT	1224
Db	2400		TCAGTGTAAACTAGAACAGCTATTTCAGGAGCACCGTTTGGTCTCTCACTATAACACTTCT	2459
Qy	1225		CAGAGATGCTATATTTCTGTGAAACACTGAACCTCGCTCTCTCCAGATTAAGCAAAAAGG	1284
Db	2460		CAGAGATGCTATATTTCTGTGAAACACTGAACCTCGCTCTCTCCAGATTAAGCAAAAAGG	2519
Qy	1285		AGCAAAAACAGACTTTTCAAGAAATGATGAATTTACATTCAGATCTGTTTAGTCAAGTGTAT	1344
Db	2520		AGCAAAAACAGACTTTTCAAGAAATGATGAATTTACATTCAGATCTGTTTAGTCAAGTGTAT	2579
Qy	1345		TGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGAATGGCTTTACAGCAACCACT	1404
Db	2580		TGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGAATGGCTTTACAGCAACCACT	2639
Qy	1405		ACTCAACAAGCAGCTGACTTATGTGTTTATTTGGAATTTGGATTCAGAACTGTTTCCAGA	1464
Db	2640		ACTCAACAAGCAGCTGACTTATGTGTTTATTTGGAATTTGGATTCAGAACTGTTTCCAGA	2699
Qy	1465		GCTCAATAAGGTAACAAGGAAGTTACTCTGTGCACATCTTGGATGTAAACACTTGGATT	1524
Db	2700		GCTCAATAAGGTAACAAGGAAGTTACTCTGTGCACATCTTGGATGTAAACACTTGGATT	2759
Qy	1525		TGGTATAGAATAAACCCATTGAAAATTTCTGCTGTCGGAGGTGGTAGAAAATTTACTTTTTT	1584
Db	2760		TGGTATAGAATAAACCCATTGAAAATTTCTGCTGTCGGAGGTGGTAGAAAATTTACTTTTTT	2819
Qy	1585		GGGTATATCTTATATATATATGTAATCATCGCTCTGTGAAAATTTTAGTATTTTTTCTTT	1644

Db 2820 GGGTATATTCCTTATATATATATGATACATCGCTGCTCGAAATTTTAGTTATTTTTCGTTT 287

Qy 1645 TTAATAA 1651  
|||||

Db 2880 TTAATAA 2886

RESULT 6  
AK095380 3616 bp mRNA linear PRI 15-JUL-2002  
LOCUS Homo sapiens cDNA FLJ38061 fis, clone CTONG2014966, highly similar  
DEFINITION to SORTING NEXIN 14.  
ACCESSION AK095380  
VERSION AK095380.1 GI:21754626  
KEYWORDS: oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE 1  
AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M., Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshina,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuhio,Y., Nagai,K. and Tsogai,T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3616)  
AUTHORS Isogai,T. and Yamamoto,J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB.  
FEATURES  
source Location/Qualifiers  
1. 3616  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CTONG2014966"  
/tissue\_type="tongue, tumor tissue"  
/clone\_lib="CTONG2"  
/note="Cloning vector: pME18SFL3"  
BASE COUNT 1078 a 629 c 710 g 1199 t  
ORIGIN  
Query Match 66.0%; Score 1315; DB 9; Length 3616;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1535; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
Qy 145 CAGGAAACACAGAAAAAGGGAGGAATCATTTGGGAATCAGCAGAAATAGGTAGCAAAATTA 204  
Db 2079 CAGGAAACACAGAAAAAGGGAGGAATCATTTGGGAATCAGCAGAAATAGGTAGCAAAATTA 2138  
Qy 205 AGGAGTATTCAAAAGTACCAATGAGGGAGGCTATGTTGCCTAATTTAGGTAGCTGA 264  
Db 2139 AGGAGTATTCAAAAGTACCAATGAGGGAGGCTATGTTGCCTAATTTAGGTAGCTGA 2198  
Qy 265 AGGTGAAGATGATTTTATTCGAAGAAGGTATTTGTTGAATGGAAGATCATTTCTCCAGTGA 324  
Db 2199 AGGTGAAGATGATTTTATTCGAAGAGGTATTTGTTGAATGGAAGATCATTTCTCCAGTGA 2258  
Qy 325 GCCTGTGAGCACACCTAATACTACCCCGAAACCTTGCCTGCATGGAAATTTAGCATTCCTA 384

Db 2259 GGCTGTGAGCACACCTTAATACCTCCCGAAAACTTGTGTCATGGAATAATTAGCATTTCCATA 2318  
Qy TGTAGACTTTTGTGAGTCCCTCTCTGAGAGGAGGAG-AAAAAGAAAGAAATTCCTG 443  
Db 2319 TGTAGACTTTTGTGAGTCCCTCTCTGAGAGGAGGAG-AAAAAGAAAGAAATTCCTG 2378  
Qy TGTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACAGCCCTGACACTT 503  
Db 2379 TGTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACAGCCCTGACACTT 2438  
Qy 504 GGTCTGTATAGAGATCTTTGAAATCTATGATCTTGAATCAAACTAAACAGAAATTC 563  
Db 2439 GGTCTGTATAGAGATCTTTGAAATCTATGATCTTGAATCAAACTAAACAGAAATTC 2498  
Qy 564 ATGGTGCATTTCTGTATGCCAGCTTCCCTCTTAAGAGGATCATTTGGCCCCCAAAATATG 623  
Db 2499 ATGGTGCATTTCTGTATGCCAGCTTCCCTCTTAAGAGGATCATTTGGCCCCCAAAATATG 2558  
Qy 624 AATTTCTTAAAGTCAAGAGGAGAGATTTCAAGATATCTACAGAACTTTCTGCAGCATC 683  
Db 2559 AATTTCTTAAAGTCAAGAGGAGAGATTTCAAGATATCTACAGAACTTTCTGCAGCATC 2618  
Qy 684 CAGAACTGAGTAAATAGTCAACTTTCTGCAGACTTTCTTTCCCTAAATGTTGGGAAACAC 743  
Db 2619 CAGAACTGAGTAAATAGTCAACTTTCTGCAGACTTTCTTTCCCTAAATGTTGGGAAACAC 2678  
Qy 744 AATTTCTTGAATAGATCTACAGATGTAATCTTGGGAAATTAATAATCTGTTCTG 803  
Db 2679 AATTTCTTGAATAGATCTACAGATGTAATCTTGGGAAATTAATAATCTGTTCTG 2738  
Qy 804 GAAACTTAATGAAGAGAGAGTCAAGATTTGAACTTTTATCATGAATTTCAATTAAT 863  
Db 2739 GAAACTTAATGAAGAGAGAGTCAAGATTTGAACTTTTATCATGAATTTCAATTAAT 2798  
Qy 864 CTGTGTAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTT 923  
Db 2799 CTGTGTAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTT 2858  
Qy 924 CAGAAACACACAGAGCTTTCAATGATCTGTTTAAATAATGCAAAACCGTGTGAAA 983  
Db 2859 CAGAAACACACAGAGCTTTCAATGATCTGTTTAAATAATGCAAAACCGTGTGAAA 2918  
Qy 984 ATACAGAGAGAGAGAGAGTCAAGATTTTATGAGGAGTATGATCTGAGAGGAGTCT 1043  
Db 2919 ATACAGAGAGAGAGAGTCAAGATTTTATGAGGAGTATGATCTGAGAGGAGTCT 2978  
Qy 1044 ATGATTACCTGATCTATGTAGAGCGGTAGTTTTCAGAGTCTCTGAGTGGCTTCATCATC 1103  
Db 2979 ATGATTACCTGATCTATGTAGAGCGGTAGTTTTCAGAGTCTCTGAGTGGCTTCATCATC 3038  
Qy 1104 TCTTAAATGGGAACTCGAATCTCTTTTAAACACACCTTGGAAATGTATGATTAATCATC 1163  
Db 3039 TCTTAAATGGGAACTCGAATCTCTTTTAAACACACCTTGGAAATGTATGATTAATCATC 3098  
Qy 1164 TTGAGTGAAGTGAAGTCAAGTATTTTCAAGAGAGCGGTGGTCTCACTATTAACACTTC 1223  
Db 3099 TTGAGTGAAGTGAAGTCAAGTATTTTCAAGAGAGCGGTGGTCTCACTATTAACACTTC 3158  
Qy 1224 TCAGAGATGCTATATCTGTGAAACACACTGAACTCTCTCCAGAGTAAAGCAAAAG 1283  
Db 3159 TCAGAGATGCTATATCTGTGAAACACACTGAACTCTCTCCAGAGTAAAGCAAAAG 3218  
Qy 1284 GAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGA 1343  
Db 3219 GAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGA 3278  
Qy 1344 TTGAGTGAAGTGAAGTCAAGTATGAAAGCATCAGACTTTCTGTTGATGCTTACAGCAACAG 1403  
Db 3279 TTGAGTGAAGTGAAGTCAAGTATGAAAGCATCAGACTTTCTGTTGATGCTTACAGCAACAG 3338  
Qy 1404 TACTCAACAGAGTCACTTATGTTTATTTGACATTTGATACAGAACTGTTTCCAG 1463

Db 3339 TACTCAACAGAGAGCTGACTTATGTTTATTTGACATTTGATGATACAGGAACTGTTTCCAG 3398  
Qy 1464 AGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAAACAATTGGAT 1523  
Db 3399 AGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAAACAATTGGAT 3458  
Qy 1524 TTGGTATAGAAATTAACCCATTTGAATTTCTGCTGTGCGAGGTTGTAGAAATTTACTTTT 1583  
Db 3459 TTGGTATAGAAATTAACCCATTTGAATTTCTGCTGTGCGAGGTTGTAGAAATTTACTTTT 3518  
Qy 1584 TTGGTATATTTCTTATATATATATATGATGATCATGCTGCTGCTGAAATTTTATTTTGT 1643  
Db 3519 TTGGTATATTTCTTATATATATATATGATGATCATGCTGCTGCTGAAATTTTATTTTGT 3578  
Qy 1644 TTTAATAAGAGACTAAACAACTTAATGATTTAAAGTG 1681  
Db 3579 TTTAATAAGAGACTAAACAACTTAATGATTTAAAGTG 3616

RESULT 7  
AX512835  
LOCUS AX512835 3145 bp DNA linear PAT 03-OCT-2002  
DEFINITION Sequence 4 from Patent WO02062839.  
ACCESSION AX512835  
VERSION AX512835.1 GI:23504019  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Daemen, M.J., Cleutjens, C.B. and Zaman, G.J.  
Markers of unstable atherosclerotic plaques  
Patent: WO 02062839-A 4 15-AUG-2002;  
Universiteit Maastricht (NL)  
Location/Qualifiers  
1. 3145  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 985 a 570 c 659 g 931 t  
ORIGIN

Query Match 65.1%; Score 1296; DB 6; Length 3145;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1446; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 147 GGAACACACAGAAAGGGGAGATCAATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 206  
Db 1602 GGAACACACAGAAAGGGGAGATCAATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 1661  
Qy 207 GAGTATTCAAAAGTACCAATGAGGAGCTATTTGCCCTAATATGTTAGTGTAGCTGAG 266  
Db 1662 GAGTATTCAAAAGTACCAATGAGGAGCTATTTGCCCTAATATGTTAGTGTAGCTGAG 1721  
Qy 267 GTGAGATGATTTTATGAAAGGATTTGTTGAAATGGAAGATGATTTCCAGTGAGG 326  
Db 1722 GTGAGATGATTTTATGAAAGGATTTGTTGAAATGGAAGATGATTTCCAGTGAGG 1781  
Qy 327 CTGTGAGCACACCTAAATACTCCCGAAACCTTGTGCTGATGGAATAATAGCAATTCATG 386  
Db 1782 CTGTGAGCACACCTAAATACTCCCGAAACCTTGTGCTGATGGAATAATAGCAATTCATG 1841  
Qy 387 TAGACTTTTTTGAAGATCCCTCTCTGAAAGAGAGGAGAAAGAAAGAAATTCCTGTGT 446  
Db 1842 TAGACTTTTTTGAAGATCCCTCTCTGAAAGAGAGGAGAAAGAAAGAAATTCCTGTGT 1901  
Qy 447 TTTGATTTGATTTGAAAGAAATGATAGAGAGAGTTGGACAGCCTGAAACATTTGT 506  
Db 1902 TTTGATTTGATTTGAAAGAAATGATAGAGAGAGTTGGACAGCCTGAAACATTTGT 1961  
Qy 507 CTGTCTATAGAGATATCTGAAATTCATGTTGATTTGAATCAAACTAACGAAATTCATG 566

Db 1962 CTGCTATAGAGATATCTTGAATTCATGTAATCTGAATCAAACTAACAGAAATTTTCATG 2021  
Qy 567 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAAT 626  
Db 2022 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAAT 2081  
Qy 627 TCTTAAAGTCAAGAGAGGAGAGTTCACAGATATCTACAGAAACTTCTGCGACATCCAG 686  
Db 2082 TCTTAAAGTCAAGAGAGGAGAGTTCACAGATATCTACAGAAACTTCTGCGACATCCAG 2141  
Qy 687 AACTGAGTAATAGTCAACTTCTGCGAGACTTCTTCCCTTAAATGTTGGGGAACACAAAT 746  
Db 2142 AACTGAGTAATAGTCAACTTCTGCGAGACTTCTTCCCTTAAATGTTGGGGAACACAAAT 2201  
Qy 747 TCTTGAATAGTACACAGATGTAATCTTGGGAATTAATAATCTGTTCTCGGAA 806  
Db 2202 TCTTGAATAGTACACAGATGTAATCTTGGGAATTAATAATCTGTTCTCGGAA 2261  
Qy 807 AACTGAATGAAGAGAGAGTTCAGCATTTGGAACCTTTTATCATGAATTTCAATATCTT 866  
Db 2262 AACTGAATGAAGAGAGAGTTCAGCATTTGGAACCTTTTATCATGAATTTCAATATCTT 2321  
Qy 867 GTGAGTCTCAAGAGCTTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 926  
Db 2322 GTGAGTCTCAAGAGCTTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 2381  
Qy 927 AAACACACAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGGAATA 986  
Db 2382 AAACACACAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGGAATA 2441  
Qy 987 CAGAGAGAGAGAGAGAGTTCAGATTTATGAGAGTGAATGAGTGAAGAGAGTCTATG 1046  
Db 2442 CAGAGAGAGAGAGAGTTCAGATTTATGAGAGTGAATGAGTGAAGAGAGTCTATG 2501  
Qy 1047 ATTACCTGATGATAGGAGCGGTAGTTTTCAGGTTCTCTGAGTGGCTTCATCTCTCT 1106  
Db 2502 ATTACCTGATGATAGGAGCGGTAGTTTTCAGGTTCTCTGAGTGGCTTCATCTCTCT 2561  
Qy 1107 TAATGGGAATCGAATCTCTTTTAAACACACCTCGGAATGATATGATTAATCTATCTTC 1166  
Db 2562 TAATGGGAATCGAATCTCTTTTAAACACACCTCGGAATGATATGATTAATCTATCTTC 2621  
Qy 1167 AGTGTAACTAGACAGCTATTTTCAGAGACACCGTTTGGTCTCACTCATAACTCTCA 1226  
Db 2622 AGTGTAACTAGACAGCTATTTTCAGAGACACCGTTTGGTCTCACTCATAACTCTCA 2681  
Qy 1227 GAGATGCTATTTCTGGAACACACTGAACTCGCTCTCTCCAGATTAAGCAAAAGGAG 1286  
Db 2682 GAGATGCTATTTCTGGAACACACTGAACTCGCTCTCTCCAGATTAAGCAAAAGGAG 2741  
Qy 1287 CAACACAGACTTTTGAAGAAATGATGAATTCATTCAGATCTGTTAGTCAAGTGATTTG 1346  
Db 2742 CAACACAGACTTTTGAAGAAATGATGAATTCATTCAGATCTGTTAGTCAAGTGATTTG 2801  
Qy 1347 GTGAGAAACCAAGTATGAAGCATCGACTTCTGTTTGAATGAGCTTACAGCAACGATAC 1406  
Db 2802 GTGAGAAACCAAGTATGAAGCATCGACTTCTGTTTGAATGAGCTTACAGCAACGATAC 2861  
Qy 1407 TCACAGAGAGCTGACTTATGTTTATGACATTTGATACAGGAACTGTTTCCAGAGC 1466  
Db 2862 TCACAGAGAGCTGACTTATGTTTATGACATTTGATACAGGAACTGTTTCCAGAGC 2921  
Qy 1467 TCATATAGGTACAAAGAGAGTTCACCTCTGTCATCTTGGATGTAACACTTGGATTG 1526  
Db 2922 TCATATAGGTACAAAGAGAGTTCACCTCTGTCATCTTGGATGTAACACTTGGATTG 2981  
Qy 1527 GTATAGAAATACCCATTGAAATTTCTGCTGCGAGGGTGTAGAAATTTACTTTTTGG 1586  
Db 2982 GTATAGAAATACCCATTGAAATTTCTGCTGCGAGGGTGTAGAAATTTACTTTTTGG 3041  
Qy 1587 GTATATTTCT 1595  
Db 3042 GTATATTTCT 3050

## RESULT 8

AK000362

## LOCUS

## DEFINITION

AK000362

## ACCESSION

AK000362

## VERSION

AK000362.1

## KEYWORDS

oligo capping, fis (full insert sequence).

## SOURCE

Homo sapiens

## ORGANISM

Homo sapiens

## REFERENCE

1 (sites)

## AUTHORS

Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,

Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,

Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,

Nakamura, Y., Isogai, T. and Sugano, S.

## TITLE

NEDO human cDNA sequencing project

## JOURNAL

Unpublished

## AUTHORS

2 (bases 1 to 3145)

## TITLE

Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,

Shibahara, T., Tanaka, T. and Nakamura, Y.

## JOURNAL

Direct Submission

Submitted (15-FEB-2000)

Sumio Sugano, Institute of Medical Science,

University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,

Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp,

Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)

## COMMENT

NEDO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan; cDNA full insert

sequencing; Research Association for Biotechnology; cDNA library

construction, 5'- &amp; 3'-end one pass sequencing; Department of

Virology and Human Genome Center, Institute of Medical Science,

University of Tokyo (partly supported by Science and Technology

Agency).

## FEATURES

Location/Qualifiers

1..3145

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="HEP15804"

/cell\_line="HepG2"

/cell\_type="hepatoma"

/clone\_lib="HEP"

/note="cloning vector pME18SFL3"

misc\_feature

1..3145

/note="highly similar to AF121863 Homo sapiens sorting

nexin 14"

BASE COUNT

985 a 570 c 659 g 931 t

ORIGIN

Query Match 65.1%; Score 1296; DB 9; Length 3145;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1446; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 147 GGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 206

Db 1602 GGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 1661

Qy 207 GAGTATTCAGAAAGTACCAATGGAGGAGCTATGTTGCCTTAATATTGGTGTAGCTGAG 266

Db 1662 GAGTATTCAGAAAGTACCAATGGAGGAGCTATGTTGCCTTAATATTGGTGTAGCTGAG 1721

Qy 267 GTGAAGATGATTTTATGAAGAGGATTTGTTGAATGGAAGATGATTTCCAGTGGAGG 326

Db 1722 GTGAAGATGATTTTATGAAGAGGATTTGTTGAATGGAAGATGATTTCCAGTGGAGG 1781

Qy 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTGTGCTGATCGAAATTAAGCAATTCATATG 386

Db 1782 CTGTGAGCACACCTTAATCTCCCGAAACCTTGTGCTGATCGAAATTAAGCAATTCATATG 1841

Qy 387 TAGACTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAG 446



QY 265 AGGTGAAGATGATTTTATTTGAGAGAGGTATGTTGTAATGGAAGATGATCTCCAGTGA 324  
Db 2136 AGGTGAAGATGATTTTATTTGAGAGAGGTATGTTGTAATGGAAGATGATCTCCAGTGA 2195  
QY 325 GGCTGTGAGACACCTTAATCTCCCGAAACCTTGCTGCATCGAATAATTTAGCATTTCCATA 384  
Db 2196 GGCTGTGAGACACCTTAATCTCCCGAAACCTTGCTGCATCGAATAATTTAGCATTTCCATA 2255  
QY 385 TGTAGACTTTTTTGTAGGATCCCTCTCTGAAAGGAGAGAGAAAAAGAAATTCCTGT 444  
Db 2256 TGTAGACTTTTTTGTAGGATCCCTCTCTGAAAGGAGAGAGAAAAAGAAATTCCTGT 2315  
QY 445 GTTTGTATGATGTTGAAAGAAATGATAGAGAGAGTTGGACACAGCCTGGAACATTTG 504  
Db 2316 GTTTGTATGATGTTGAAAGAAATGATAGAGAGAGTTGGACACAGCCTGGAACATTTG 2375  
QY 505 GTCTGTCTATAGAGATATCTGAATCTTATGATCTTGAATCAAACTAACAGATTTCA 564  
Db 2376 GTCTGTCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTAACAGATTTCA 2435  
QY 565 TGTGTGATTTCCGATGCCCCAGCTTCTCTTAAGAGATCATTTGGCCCCAAAAATATGA 624  
Db 2436 TGTGTGATTTCCGATGCCCCAGCTTCTCTTAAGAGATCATTTGGCCCCAAAAATATGA 2495  
QY 625 ATTCTTTAAGTCAAAGGAGGAGGATTTCCAGAAATATCTACAGAAATCTTCGAGCATCC 684  
Db 2496 ATTCTTTAAGTCAAAGGAGGAGGATTTCCAGAAATATCTACAGAAATCTTCGAGCATCC 2555  
QY 685 AGAAGTGAATAGTCACTTCTGAGAGACTTTCTTCCCTTAAGGATTTGAGGAGGAGACACA 744  
Db 2556 AGAAGTGAATAGTCACTTCTGAGAGACTTTCTTCCCTTAAGGATTTGAGGAGGAGACACA 2615  
QY 745 ATTCTTTGATAGATCTACAGATGTAATCTTGGGAAATATTAATAATCTGTTCTCGG 804  
Db 2616 ATTCTTTGATAGATCTACAGATGTAATCTTGGGAAATATTAATAATCTGTTCTCGG 2675  
QY 805 AAAAATAAGAGAGAGAGGATGAGATTTGAGACCTTTTATCATGAATTTCAATTAATTC 864  
Db 2676 AAAAATAAGAGAGAGAGGATGAGATTTGAGACCTTTTATCATGAATTTCAATTAATTC 2735  
QY 865 TTGTGATCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTGAGCCTACTTTC 924  
Db 2736 TTGTGATCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTGAGCCTACTTTC 2795  
QY 925 AGAAAAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAAA 984  
Db 2796 AGAAAAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAAA 2855  
QY 985 TACAGAGAGAGAGAGAGAGATTTTATGAGGATGATGATGATGAGAGAGGATCTA 1044  
Db 2856 TACAGAGAGAGAGAGAGAGATTTTATGAGGATGATGATGATGAGAGAGGATCTA 2915  
QY 1045 TGATTACCTGATGATGATGAGAGCGGTAGTTTCCAGGTTTCTGACTGGCTTCATCATCT 1104  
Db 2916 TGATTACCTGATGATGATGAGAGCGGTAGTTTCCAGGTTTCTGACTGGCTTCATCATCT 2975  
QY 1105 CTTAATGGGAACTCGAATCTCTTTTAAACAACCTCGGAAATGATGATGATGATCT 1164  
Db 2976 CTTAATGGGAACTCGAATCTCTTTTAAACAACCTCGGAAATGATGATGATGATCT 3035  
QY 1165 TCAGTGTAACTAGAACAGCTATTTTACGAGGACCGTTTGGTCTCACTCATACATCTT 1224  
Db 3036 TCAGTGTAACTAGAACAGCTATTTTACGAGGACCGTTTGGTCTCACTCATACATCTT 3095  
QY 1225 CAGAGATGCTATATTTCTGTGAAACACTGAACTCTGCTCTCTCAAGATGAGCAAAAGG 1284  
Db 3096 CAGAGATGCTATATTTCTGTGAAACCTGAACTCTGCTCTCTCAAGATGAGCAAAAGG 3155  
QY 1285 AGCAAAACAGACTTTTGAAGAAATGATGAATTAATTTCCAGATCTGTTAGTCAAGTGTAT 1344  
Db 3156 AGCAAAACAGACTTTTGAAGAAATGATGAATTAATTTCCAGATCTGTTAGTCAAGTGTAT 3215

QY 1345 TGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGTGGCTTACAGCAACCACT 1404  
Db 3216 TGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGTGGCTTACAGCAACCACT 3275  
QY 1405 ACTCAACAAGCAG 1417  
Db 3276 ACTCAACAAGCAG 3288  
RESULT 10  
LOCUS HSA420561 968 bp mRNA linear PRI 23-NOV-2001  
DEFINITION Homo sapiens mRNA full length insert cdna clone EUROIMAGE 1521079.  
ACCESSION AJ420561  
VERSION AJ420561.1 GI:17066425  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Auffray, C., Ansong, W., Ballabio, A., Estivill, X., Gibson, K.,  
Lehrach, H., Poustka, A. and Lundberg, J.  
TITLE The European IMAGE consortium for integrated Molecular analysis of  
human gene transcripts  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 968)  
AUTHORS Persson, A.  
TITLE Direct Submission  
JOURNAL Submitted (02-OCT-2001) Persson A., Center for Molecular  
Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagsvagen  
30B, 106 91 Stockholm, SWEDEN  
COMMENT This clone is available royalty-free through IMAGE Consortium  
Distributors. IMPORTANT: This sequence represents the full insert  
of this IMAGE cdna clone. No attempt has been made to verify  
whether this corresponds to the full-length of the original mRNA  
from which it was derived.  
FEATURES  
Location/Qualifiers  
1..968  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clones="EUROIMAGE 1521079"  
/clone\_lib="NCI CGAP Lu5"  
BASE COUNT 311 a 147 c 186 g 324 t  
ORIGIN  
Query Match 41.0%; Score 816; DB 9; Length 968;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 936; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1021 GGTGATGATCTGAGAGAGGCTTATGATTACCTGATGATGAGAGGCTGAGTTTCCA 1080  
Db 20 GGTGATGATCTGAGAGAGGCTTATGATTACCTGATGATGAGAGGCTGAGTTTCCA 79  
QY 1081 GGTCTCTGACCTGGCTTCATCATCTCTTAATGGGAACCTGAAATCTCTTTAAAAACACCT 1140  
Db 80 GGTCTCTGACCTGGCTTCATCATCTCTTAATGGGAACCTGAAATCTCTTTAAAAACACCT 139  
QY 1141 GGAATGTATATCTGATTAATCTCTTCTGATGATGATGATGATGATGATGATGATGAT 1200  
Db 140 GGAATGTATATCTGATTAATCTCTTCTGATGATGATGATGATGATGATGATGATGAT 199  
QY 1201 TTTGTTCTCATCTATAACATCTTCAGAGATGCTATATTTCTGTGAAAACACTGAACTTCG 1260  
Db 200 TTTGTTCTCATCTATAACATCTTCAGAGATGCTATATTTCTGTGAAAACACTGAACTTCG 259  
QY 1261 CTCTCTCCAGATGAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTTACAT 1320  
Db 260 CTCTCTCCAGATGAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTTACAT 319  
QY 1321 TCCAGATCTGTTAGTCAAGTGTATTGTTGTAAGAAACCAAGTATGAAAGCATCAGACTTCT 1380







```
ACCESSION BD113855
VERSION JP 2002010789-A/5932
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 425)
AUTHORS Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.
TITLE EST and encoded human protein
JOURNAL Patent: JP 2002010789-A 5932 15-JAN-2002;
GENSET CORP
COMMENT OS Homo sapiens (human)
PN JP 2002010789-A/5932
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 05-AUG-1999 US 60/147499
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EYE PI
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21,
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC EST and encoded human protein
FH Key Location/Qualifiers
FT source 1..425
FEATURES
source
1..425
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 134 a 82 c 85 g 124 t
ORIGIN
Query Match 21.38; Score 425; DB 6; Length 425;
Best Local Similarity 100.0%; Pred. No. 8.3e-201;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1004 CAGAAATATATTTATGGAGGTGATGACCTGTAGAGGAGTCTATGATTAACCTGATGATGA 1063
DB 1 CAGAAATATATTTATGGAGGTGATGACCTGTAGAGGAGTCTATGATTAACCTGATGATGA 60
QY 1064 GGACGGGTAGTTTCCAGGTTCTGACTGCTTCATCACTCTTAATGGGAACCTGAATC 1123
DB 61 GGACGGGTAGTTTCCAGGTTCTGACTGCTTCATCACTCTTAATGGGAACCTGAATC 120
QY 1124 CTCTTTAAACACCCCTGGAATGTATCTGATTAATCTTCACTGCTGATTAACCTGATGA 1183
DB 121 CTCTTTAAACACCCCTGGAATGTATCTGATTAATCTTCACTGCTGATTAACCTGATGA 180
QY 1184 CTATTTCCAGGAGCACCGTTTGGTCTCACTCATCACTCTTCACTGATGATCTATCTCT 1243
DB 181 CTATTTCCAGGAGCACCGTTTGGTCTCACTCATCACTCTTCACTGATGATCTATCTCT 240
QY 1244 GAAACACTGAACCTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGA 1303
DB 241 GAAACACTGAACCTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGA 300
QY 1304 GAAATGATGAATATCACTTCAGATCTGTTAGTCAAGTATGTTGTTGTTGTTGTTGTT 1363
DB 301 GAAATGATGAATATCACTTCAGATCTGTTAGTCAAGTATGTTGTTGTTGTTGTTGTT 360
QY 1364 GAAAGCATCAGACTCTGTTTGGCTTACGACCACTACTCTCAACAGCAGCTGACT 1423
DB 361 GAAAGCATCAGACTCTGTTTGGCTTACGACCACTACTCTCAACAGCAGCTGACT 420
QY 1424 TATGT 1428
DB 421 TATGT 425
RESULT 13
AX331384/C
LOCUS AX331384 451 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 1893 from Patent WO0194629.
ACCESSION AX331384
VERSION AX331384.1 GI:18122018
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrikan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
JOURNAL Patent: WO 0194629-A 1893 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..451
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 179 a 73 c 52 g 146 t
ORIGIN
Query Match 19.38; Score 385; DB 6; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.3e-181;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1597 ATATATATATGATCATCGCTGCTGAAATTTAGTTATTTTGTGTTTAAATAAGACT 1656
DB 385 ATATATATATGATCATCGCTGCTGAAATTTAGTTATTTTGTGTTTAAATAAGACT 326
QY 1657 AACCAAACTTATGATTTAAAGTGATGATGATGATGATGATGATGATGATGATGATG 1716
DB 325 AACCAAACTTATGATTTAAAGTGATGATGATGATGATGATGATGATGATGATGATG 266
QY 1717 TCGAAATTTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1776
DB 265 TCGAAATTTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206
QY 1777 AATGATGTTATGCGAAACAGATGATGATGATGATGATGATGATGATGATGATGATG 1836
DB 205 AATGATGTTATGCGAAACAGATGATGATGATGATGATGATGATGATGATGATGATG 146
QY 1837 ATTATGCTAGTATGAGAAACAGGAAATAGATCTGATTTTCTTAGAGTTAATATATTT 1896
DB 145 ATTATGCTAGTATGAGAAACAGGAAATAGATCTGATTTTCTTAGAGTTAATATATTT 86
QY 1897 AGTAGATTGGTTTCCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1956
DB 85 AGTAGATTGGTTTCCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 26
QY 1957 GCATCCTATATGATGATTTTATATAT 1981
DB 25 GCATCCTATATGATGATTTTATATAT 1
RESULT 14
BD026500
LOCUS BD026500 358 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD026500
VERSION BD026500.1 GI:22567723
KEYWORDS JP 2001269182-A/2746.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 358)
AUTHORS Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
TITLE Sequence tag and encoded human protein
JOURNAL Patent: JP 2001269182-A 2746 02-OCT-2001;
```



```

GENSET
OS Homo sapiens (human)
PN JP 2001269182-A/2746
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCCLAIR, JEAN YVES
PI JORDAN
PC C12N15/09.C07K14/435.C07K16/18.C12N1/15.C12N1/19.C12N1/21, PC
PC C12P21/02.C12P21/08.C12Q1/68//G06F17/30.C12N15/00.C12N5/00, PC
G06F15/40
CC Key Location/Qualifiers
FH CDS 53..358.
FT Location/Qualifiers
FEATURES
source
1..358
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 125 a 63 c 69 g 101 t
ORIGIN
Query Match 17.9%; Score 357; DB 6; Length 358;
Best Local Similarity 100.0%; Pred. No. 8.6e-167;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 760 ACTACAGATGTAATCTTGGAATAATATAAATCTGTTCTCGGAAACTAATGAAAGA 819
Db 1 ACTACAGATGTAATCTTGGAATAATATAAATCTGTTCTCGGAAACTAATGAAAGA 60
QY 820 GAAAGGTGAGATTTGGAACCTTTATCATGAAATTTCAATTAATCTTGAGTCTCCAA 879
Db 61 GAAAGGTGAGATTTGGAACCTTTATCATGAAATTTCAATTAATCTTGAGTCTCCAA 120
QY 880 GCCTAAACCAAGTAGACAGAACCTGACATCTCAGCCCTACTTCAGAAACCAACAGAA 939
Db 121 GCCTAAACCAAGTAGACAGAACCTGACATCTCAGCCCTACTTCAGAAACCAACAGAA 180
QY 940 GCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTGCTGAAATAACAGAGAGAAAGCA 999
Db 181 GCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTGCTGAAATAACAGAGAGAAAGCA 240
QY 1000 AAATCAGAAATATTTATGAGGTGATGACCTGTAGAGAGAGTCTATGATTAACCTGATGA 1059
Db 241 AAATCAGAAATATTTATGAGGTGATGACCTGTAGAGAGAGTCTATGATTAACCTGATGA 300
QY 1060 TGTAGGCGGTAGTTTCCAGGTTCTGAGTTCATCATCTCTTAATGGGAAAC 1116
Db 301 TGTAGGCGGTAGTTTCCAGGTTCTGAGTTCATCATCTCTTAATGGGAAAC 357

RESULT 15
G30543
LOCUS human STS SHGC-37370, sequence tagged site. 454 bp DNA linear STS 05-OCT-1996
DEFINITION human STS SHGC-37370, sequence tagged site.
ACCESSION G30543
VERSION G30543.1 GI:1594094
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Myers,R.M.
JOURNAL Unpublished (1996)
CONTACT: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689

```

Email: myers@shgc.stanford.edu

Primer A: ACTGACAGATTCTTCAGACTTT

Primer B: GCAATCAATACTCCCAAA

STS size: 150

PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds

Annealing: 62 degrees C for 23 seconds

Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Taq Polymerase: 0.05 units/ul

Total Vol: 10 ul

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 20 mM

pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from N48004  
-- Washington University/Merck EST sequence.

#### FEATURES

source

1..454

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/map="6"

1..150

primer bind

primer\_bind complement(131..150)

BASE COUNT 129 a 79 c 97 g 147 t 2 others

#### ORIGIN

Query Match

Best Local Similarity 7.2%; Score 143; DB 11; Length 454;

Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 AGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCGTT 1383

Db 259 AGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCGTT 318

QY 1384 TGATGGCTTACAGCAACCAAGTACTCAACAAGCAGCTGACTTATGTTTTATTGGACATTGT 1443

Db 319 TGATGGCTTACAGCAACCAAGTACTCAACAAGCAGCTGACTTATGTTTTATTGGACATTGT 378

QY 1444 GATACAGGAACCTGTTTCCAGAGC 1466

Db 379 GATACAGGAACCTGTTTCCAGAGC 401

Search completed: January 31, 2004, 18:20:55

Job time : 7423 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 31, 2004, 15:43:54 ; Search time 4299 Seconds  
(without alignments)  
11261.809 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1 gtaagaacacccaaagtt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO NUC  
Gapop 60.0, Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : EST:

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_ges\_hum:\*
- 18: em\_ges\_inv:\*
- 19: em\_ges\_pin:\*
- 20: em\_ges\_vrt:\*
- 21: em\_ges\_fun:\*
- 22: em\_ges\_mam:\*
- 23: em\_ges\_mus:\*
- 24: em\_ges\_pro:\*
- 25: em\_ges\_rod:\*
- 26: em\_ges\_pbg:\*
- 27: em\_ges\_vrl:\*
- 28: gb\_ges1:\*
- 29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	706	35.4	1096	12	BQ053127
C 2	690	34.6	702	14	CB054302 NISC Gm03
C 3	680	34.1	921	13	BQ230190 AGENCOURT
C 4	673	33.8	773	12	BM982264 UI-CF-EN1

5	652	32.7	664	14	CB136167
C 6	651	32.7	917	13	BX372281
C 7	640	32.1	708	13	BU634170
C 8	636	31.9	643	14	CB114674
9	620	31.1	1105	10	BG256631
C 10	619	31.1	802	14	CA431124
C 11	613	30.8	629	13	BU579763
C 12	608	30.5	608	12	BM829345
C 13	606	30.4	606	12	BM699448
C 14	599	30.1	604	14	CB130920
C 15	577	29.0	594	13	BU738240
C 16	571	28.7	598	13	BU683958
C 17	570	28.6	647	13	BU679457
C 18	560	28.1	986	13	BU168735
C 19	559	28.1	559	12	BM693800
20	557	28.0	608	12	BM722250
21	548	27.5	548	14	CB139915
C 22	548	27.5	564	12	BM670092
C 23	537	27.0	1076	10	BE874550
24	534	26.8	625	10	BG536529
25	529	26.6	817	10	BF692296
26	524	26.3	717	13	BU959574
27	513	25.8	513	14	CB142612
28	507	25.5	558	14	CB160118
29	506	25.4	506	12	BM835779
30	505	25.4	616	14	CB054303
C 31	503	25.3	503	9	AI283088
C 32	503	25.3	503	9	AI955217
C 33	497	24.9	1043	9	AL574298
34	494	24.8	494	10	BG547215
35	494	24.8	726	10	BG432502
C 36	489	24.5	520	9	AW149014
37	487	24.4	638	10	AW957704
C 38	483	24.2	507	14	CA314453
C 39	480	24.1	531	9	AI949494
40	479	24.0	640	9	AW327722
C 41	479	24.0	903	9	AL578668
42	475	23.8	818	10	BG191447
43	473	23.7	473	14	CB135556
C 44	472	23.7	480	9	AI127501
45	471	23.6	479	10	AW996088

ALIGNMENTS

RESULT 1  
BQ053127  
LOCUS  
DEFINITION AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711  
5', mRNA sequence.  
ACCESSION BQ053127  
VERSION BQ053127.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1096)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: L1CM2121 row: e column: 16  
High quality sequence stop: 675.

BQ053127 1096 bp mRNA linear EST 29-MAR-2002  
AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711  
5', mRNA sequence.  
BQ053127 GI:19812467  
BQ053127.1  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1096)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: L1CM2121 row: e column: 16  
High quality sequence stop: 675.

```
FEATURES
source
Location/Qualifiers
1..1096
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5934711"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 106"
/notes="Organ: blood; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."
BASE COUNT 355 a 226 g 305 t 17 others
ORIGIN
Query Match 35.4%; Score 706; DB 12; Length 1096;
Best Local Similarity 100.0%; Pred. No. 8.9e-235;
Matches 706; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 779 GGGAAAATTATAAATCTGTTCTCGGAAACTAATGAAGAGAAAGGTGAGATTGGAA 838
DB 10 GGGAAAATTATAAATCTGTTCTCGGAAACTAATGAAGAGAAAGGTGAGATTGGAA 69
QY 839 CTTTTATCATGAATTTTCAATTAATCTTGAGTCTCCAAAGCTTAACCAAGTAGACCA 898
DB 70 CTTTTATCATGAATTTTCAATTAATCTTGAGTCTCCAAAGCTTAACCAAGTAGACCA 129
QY 899 GAACGTGACCATCTTCAGCCCTACTTCAGAAAACAAAGAGAGCTTTTCAATGATCTGTTT 958
DB 130 GAACGTGACCATCTTCAGCCCTACTTCAGAAAACAAAGAGAGCTTTTCAATGATCTGTTT 189
QY 959 AAAAATAATGCAACCGTGTGAAAATACAGAGAGAAAGCAAAATCAGAAATATTTTATG 1018
DB 190 AAAAATAATGCAACCGTGTGAAAATACAGAGAGAAAGCAAAATCAGAAATATTTTATG 249
QY 1019 GAGGTGATGATGTAGAGAGCTTATGATTAATCTGATGATGTATGAGACGGGTAGTTTC 1078
DB 250 GAGGTGATGATGTAGAGAGCTTATGATTAATCTGATGATGTATGAGACGGGTAGTTTC 309
QY 1079 CAGTTCTGCTGAGTCTCATCTCTTAATGGGAACTCGAATCTCTTTTAAACACAC 1138
DB 310 CAGTTCTGCTGAGTCTCATCTCTTAATGGGAACTCGAATCTCTTTTAAACACAC 369
QY 1139 CTGGAATGTATCTGATTAATCTTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCAC 1198
DB 370 CTGGAATGTATCTGATTAATCTTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCAC 429
QY 1199 CGTTTGGTCTCATAACACTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACCT 1258
DB 430 CGTTTGGTCTCATAACACTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACCT 489
QY 1259 CGCTCTCTCCAGATAGCAAAAGAGCAAAACAGACTTTTCAAGAAATGATGAATTAC 1318
DB 490 CGCTCTCTCCAGATAGCAAAAGAGCAAAACAGACTTTTCAAGAAATGATGAATTAC 549
QY 1319 ATTCAGATCTGTTAGTCAAGTGTATTGTTGGAACCAACCAAGTATGAAGCATCAGACTT 1378
DB 550 ATTCAGATCTGTTAGTCAAGTGTATTGTTGGAACCAACCAAGTATGAAGCATCAGACTT 609
QY 1379 CTGTTTGAATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTGGAC 1438
DB 610 CTGTTTGAATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTGGAC 669
QY 1439 ATTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGG 1484
DB 670 ATTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGG 715
```

```
RESULT 2
CB054302/c
LOCUS
DEFINITION
NISC_gm03e09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone
IMAGE:3291376 3', mRNA sequence.
ACCESSION
CB054302
VERSION
CB054302.1 GI:27792589
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 702)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM8060 row: J column: 17
Seq primer: -21M13 forward primer (ABI).
FEATURES
Location/Qualifiers
1..702
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3291376"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Brn23"
/notes="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTGAAGTGGAGGCGCGCATATCTTTTGTGTTTGTGTTTGTGTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73-vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 253 a 121 c 86 g 242 t
ORIGIN
Query Match 34.6%; Score 690; DB 14; Length 702;
Best Local Similarity 100.0%; Pred. No. 4.2e-229;
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1303 AGAATGATGAATTAATTCAGATCTGTAGTCAAGTGTATGTCGAGAAACCAAGTA 1362
DB 702 AGAATGATGAATTAATTCAGATCTGTAGTCAAGTGTATGTCGAGAAACCAAGTA 643
QY 1363 TGAAGCATCAGACTTCTGTTGATGCTTACACCAACCAAGTACTCAACAGCAGCTGAC 1422
DB 642 TGAAGCATCAGACTTCTGTTGATGCTTACACCAACCAAGTACTCAACAGCAGCTGAC 583
QY 1423 TTATGTTTATTTGACATTTGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAA 1482
DB 582 TTATGTTTATTTGACATTTGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAA 523
QY 1483 GGAAGTTACCTCTGTGACATCTTGATGTAACACTTGGATTTGGTATGATTAACCAT 1542
DB 522 GGAAGTTACCTCTGTGACATCTTGATGTAACACTTGGATTTGGTATGATTAACCAT 463
QY 1543 TGAATTTCTCTGTGAGGCTGTAGAAATTTACTTTTTTGGGTATTTCTTATATAT 1602
DB 462 TGAATTTCTCTGTGAGGCTGTAGAAATTTACTTTTTTGGGTATTTCTTATATAT 403
```



University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171

Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clones Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA-Yes.

FEATURES  
source  
Location/Qualifiers  
1. .773  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-acr-j-04-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-EN1"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-EN1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Cystic Fibrosis  
Epithelial Cells. The library was constructed according to  
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pT73-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CTGCTCAGGT.  
TAG LIB=UI-CF-EN1  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGGT"  
233 a 146 c 124 g 269 t 1 others

Query Match 33.8%; Score 673; DB 12; Length 773;  
Best Local Similarity 100.0%; Pred. No. 3.1e-223; Mismatches 0; Indels 0; Gaps 0;  
Matches 673; Conservative 0;  
Qy 1010 TATTTATGAGGTGATGACTGTAGAGGAGTCTATGATTAATGAGGAACTGAAATCTCTTT 1129  
Db 690 TATTTATGAGGTGATGACTGTAGAGGAGTCTATGATTAATGAGGAACTGAAATCTCTTT 631  
Qy 1070 GTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAATGGGAACTGAAATCTCTTT 1129  
Db 630 GTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAATGGGAACTGAAATCTCTTT 571  
Qy 1130 AAAAACCCTCGAAATGATGATGATTAATGAGGAGTCTATGATTAATGAGGAACTGAAATCTCTTT 1189  
Db 570 AAAAACCCTCGAAATGATGATGATTAATGAGGAGTCTATGATTAATGAGGAACTGAAATCTCTTT 511  
Qy 1190 CAGGAGCAGCGTTGGTCTCACTCATACACTTCTCAGAGATCTATATCTGTGAAAC 1249  
Db 510 CAGGAGCAGCGTTGGTCTCACTCATACACTTCTCAGAGATCTATATCTGTGAAAC 451  
Qy 1250 ACTGAACCTCGCTCTCTCCAGTAAGCAAAAAGGAGCAAAAACAGACTTTTGAAGAAATG 1309  
Db 450 ACTGAACCTCGCTCTCTCCAGTAAGCAAAAAGGAGCAAAAACAGACTTTTGAAGAAATG 391  
Qy 1310 ATGAATTACATCCAGATCTGTAGTCAAGTGTATGGTGAAGAAACCAAGATGAAGGC 1369  
Db 390 ATGAATTACATCCAGATCTGTAGTCAAGTGTATGGTGAAGAAACCAAGATGAAGGC 331

Qy 1370 ATCAGACTCTCTGTTGATGGCTTTACAGCACAGTACTCAACAGCAGCTGACTTATGTT 1429  
Db 330 ATCAGACTCTCTGTTGATGGCTTTACAGCACAGTACTCAACAGCAGCTGACTTATGTT 271  
Qy 1430 TTATTTGACATCTGTGATACAGGAACCTGTTTCCAGAGCTCAATTAAGGTACAAAAGGAAGTT 1489  
Db 270 TTATTTGACATCTGTGATACAGGAACCTGTTTCCAGAGCTCAATTAAGGTACAAAAGGAAGTT 211  
Qy 1490 ACCTCTGTGACATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1549  
Db 210 ACCTCTGTGACATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 151  
Qy 1550 TCTGCTGTGAGGAGGTGTAGAAATTTACTTTTTTGGTATATTTCTTTATATATATATATGTT 1609  
Db 150 TCTGCTGTGAGGAGGTGTAGAAATTTACTTTTTTGGTATATTTCTTTATATATATATATGTT 91  
Qy 1610 ACATCGCTGTCTGAATTTTAGTATTTTGTGTTTAAATAAGACTAACACAACTTAA 1669  
Db 90 ACATCGCTGTCTGAATTTTAGTATTTTGTGTTTAAATAAGACTAACACAACTTAA 31  
Qy 1670 TGATTAAGTGA 1682  
Db 30 TGATTAAGTGA 18

RESULT 5  
CB136167  
LOCUS  
DEFINITION K-EST0188577 L14ChoICK0 Homo sapiens cDNA clone L14ChoICK0-20-H05  
5' mRNA sequence.  
ACCESSION CB136167  
VERSION CB136167.1 GI:28103366  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 664)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 20 row: H column: 05  
High quality sequence stop: 664.  
Location/Qualifiers  
1. .664  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L14ChoICK0-20-H05"  
/sex="M"  
/cell\_line="Choi-CK"  
/lab\_host="Top10F"  
/clone\_lib="L14ChoICK0"  
/note="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;  
Site 2: NotI. The library was constructed by the Soares  
laboratory and it was constructed as described by Bonaldi,  
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research  
6(9): 791-806. RNA was prepared from harvested cell  
culture."

FEATURES  
source

BASE COUNT 227 a 107 c 146 g 184 t

ORIGIN

Query Match 32.7%; Score 652; DB 14; Length 664;

Best Local Similarity 100.0%; Pred. No. 6.6e-216;  
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 CACAGAAAGGGAGAGATCATTTGGAATCAGCAGAGATAGTAAAGGAGAT 212  
Db 1 CACAGAAAGGGAGAGATCATTTGGAATCAGCAGAGATAGTAAAGGAGAT 60

QY 213 TCAAAAGTACCAATGAGAGAGATATTTGCTTAATATGCTAGTGAAGTGAAG 272  
Db 61 TCAAAAGTACCAATGAGAGAGATATTTGCTTAATATGCTAGTGAAGTGAAG 120

QY 273 ATGATTTATGAAGAGATATTTGCTTAATATGCTAGTGAAGTGAAGTGAAG 332  
Db 121 ATGATTTATGAAGAGATATTTGCTTAATATGCTAGTGAAGTGAAGTGAAG 180

QY 333 GCACACTATATCTCCCAACCTTCTGCTGCAATGGAATATGCTATGAGACT 392  
Db 181 GCACACTATATCTCCCAACCTTCTGCTGCAATGGAATATGCTATGAGACT 240

QY 393 TTTTTCAGGATCCCTCTCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 452  
Db 241 TTTTTCAGGATCCCTCTCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

QY 453 TTGATCTTGAAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512  
Db 301 TTGATCTTGAAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

QY 513 ATAGAAGATATCTGAAATCTATGATGATGATGATGATGATGATGATGATG 572  
Db 361 ATAGAAGATATCTGAAATCTATGATGATGATGATGATGATGATGATGATG 420

QY 573 TTCTGATGCCAGCTTCTCTTACAGGATCATTTGGCCCAAGAGAGAGAGAGAG 632  
Db 421 TTCTGATGCCAGCTTCTCTTACAGGATCATTTGGCCCAAGAGAGAGAGAGAG 480

QY 633 AGTCAAG 692  
Db 481 AGTCAAG 540

QY 693 GTAATAGTCAACTTCTGGCAGACTTTCTTCCCTTAATGCTGGGAGAGAGAG 752  
Db 541 GTAATAGTCAACTTCTGGCAGACTTTCTTCCCTTAATGCTGGGAGAGAGAG 600

QY 753 ATAAGTACTACAGATGTAATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAG 804  
Db 601 ATAAGTACTACAGATGTAATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAG 652

RESULT 6  
BX372281/c  
LOCUS  
DEFINITION  
BX372281 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
Homo sapiens cDNA clone CS0DL007YE10 3-PRIME, mRNA sequence.  
ACCESSION  
BX372281  
VERSION  
BX372281.1 GI:30441965  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 917)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 3600.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAI0442B11\_CS04210\_1&cluster=3600.r.

Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BAI0442B11\_CS04210\_1.

FEATURES  
source  
1..917  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL007YE10"  
/cell\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 257 a 168 c 177 g 310 t  
ORIGIN

Query Match 32.7%; Score 651; DB 13; Length 917;  
Best Local Similarity 99.9%; Pred. No. 1.2e-215;  
Matches 701; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 655 AGAATATCTACAGAACTTCTGCGAGCATCCAGAACTGAGTAATAGTCAACTTCTGGCAGA 714  
Db 778 AGAATATCTACAGAACTTCTGCGAGCATCCAGAACTGAGTAATAGTCAACTTCTGGCAGA 719

QY 715 CTTTCTTTCCCTTAATGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774  
Db 718 CTTTCTTTCCCTTAATGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659

QY 775 TCTTGGGAG 834  
Db 658 TCTTGGGAG 599

QY 835 GGAACCTTTTATCATGAATTTCTTCTGAGTCTCCAAAGCTTAAACCAAGTAG 894  
Db 598 GGAACCTTTTATCATGAATTTCTTCTGAGTCTCCAAAGCTTAAACCAAGTAG 539

QY 895 ACCAGAACTACCACTTCTAGCCCTACTTCCAGAAACCAACAGAGAGTTTCAATGATCT 954  
Db 538 ACCAGAACTACCACTTCTAGCCCTACTTCCAGAAACCAACAGAGAGTTTCAATGATCT 479

QY 955 GTTTAAAAATATGCAACCCGCTGCTGAAATACAGAGAGAGAGAGAGAGAGAGAGAG 1014  
Db 478 GTTTAAAAATATGCAACCCGCTGCTGAAATACAGAGAGAGAGAGAGAGAGAGAGAG 419

QY 1015 TATCGAGTGTGATGCTAGAGAGGCTTATGATTACCTGATGTAGGACGGGTAGT 1074  
Db 418 TATCGAGTGTGATGCTAGAGAGGCTTATGATTACCTGATGTAGGACGGGTAGT 359

QY 1075 TTTCCAGGTTCTGAGCTGGCTTTCATCTCTTTAATGGGAACCTCGAATCCTCTTTAAAAA 1134  
Db 358 TTTCCAGGTTCTGAGCTGGCTTTCATCTCTTTAATGGGAACCTCGAATCCTCTTTAAAAA 299

QY 1135 CACCTCGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1194  
Db 298 CACCTCGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 239

QY 1195 GCACGTTTGGTCTCAGTCTACATACACTTCTCAGAGATGCTATTTCTGTGAAACACTGA 1254  
Db 238 GCACGTTTGGTCTCAGTCTACATACACTTCTCAGAGATGCTATTTCTGTGAAACACTGA 179

QY 1255 ACCTCGCTCTCTCAAGATAAGCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1314  
Db 178 ACCTCGCTCTCTCAAGATAAGCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119

QY 1315 TTACATTTCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAAC 1356  
Db 118 TTACATTTCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAAC 77

```

RESULT 7
BU634170/c
LOCUS      BU634170      708 bp      mRNA      linear      EST 23-SEP-2002
DEFINITION UI-H-FLI-bgw-1-08-0-UI.s1 NCI_CGAP_FLI Homo sapiens cDNA clone
VERSION     UI-H-FLI-bgw-1-08-0-UI.3', mRNA sequence.
KEYWORDS    EST.
SOURCE      BU634170.1 GI:23301425
            Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 708)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=ies.

FEATURES             Location/Qualifiers
     source           1..708
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FLI-bgw-1-08-0-UI"
                     /tissue_type="Cell lines"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI CGAP FLI"
                     /note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia)
                     with a modified polylinker; Site 1: Ecor I; Site 2: Not
                     I; NCI CGAP FLI is a normalized cDNA library derived from
                     a pool of mRNA obtained from 4 cell lines from grade III
                     chondrosarcoma tissues. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an Ecor I adaptor, digested
                     with Not I, and cloned directionally into pT73-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (drl)8 tail. The
                     sequence tag for this library is GAGGTCGGTG. The cell
                     lines were provided by Dr. James Martin from the
                     University of Iowa.
                     TAG LIB=UI-H-FLI
                     TAG_TISSUE=Human Chondrosarcoma Grade 3 cell line mix
                     TAG_SEQ=GAGGTCGGTG"

BASE COUNT      219 a 135 c 114 g 239 t 1 others.
ORIGIN
992 AGAAGCAAAATCAGAAATTTATGAGGTCGATGCTAGAGGAGTCTATGATTAC 1051
708 AGAAGCAAAATCAGAAATTTATGAGGTCGATGCTAGAGGAGTCTATGATTAC 649
1052 CTGATGTATGAGACGGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATG 1111
648 CTGATGTATGAGACGGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATG 589
1112 GGAATCGAATCTCTTTAAACACCCCTGGAATGTATCTGATTAATCTTCTAGTGT 1171

```

```

588 GGAATCGAATCTCTTTAAACACCCCTGGAATGTATCTGATTAATCTTCTAGTGT 529
1172 AAATAGACAGCTATTTTCAGGAGGACCGTTTGGTCTCTCACTCTATAACACTTCTCAGAGAT 1231
528 AAATAGACAGCTATTTTCAGGAGGACCGTTTGGTCTCTCACTCTATAACACTTCTCAGAGAT 469
1232 GCTATATTTCTGTGAAACACTGAACCTCGCTCTCTCCAGATGAAGCAAAAGGAGCAAAA 1291
468 GCTATATTTCTGTGAAACACTGAACCTCGCTCTCTCCAGATGAAGCAAAAGGAGCAAAA 409
1292 CAGACTTTTGAAGAAATGATGAATTAATCAATTCAGATCTGTTAGTCAAGTGTATTGTGAA 1351
408 CAGACTTTTGAAGAAATGATGAATTAATCAATTCAGATCTGTTAGTCAAGTGTATTGTGAA 349
1352 GAAACCAAGTATGAAGCATCAACATCTCTGTTTGGTGTTCAGCAACACGACTACTCAAC 1411
348 GAAACCAAGTATGAAGCATCAACATCTCTGTTTGGTGTTCAGCAACACGACTACTCAAC 289
1412 AAGCAGCTGACTTATGTTTATTTGGACATTTGTGATCAGGAATGTTTCCAGAGCTCAAT 1471
288 AAGCAGCTGACTTATGTTTATTTGGACATTTGTGATCAGGAATGTTTCCAGAGCTCAAT 229
1472 AAGTACAAAAGGAAGTTACTCTCTGTGACATCTTGGATGTAAACACTTGGATTGTGATA 1531
228 AAGTACAAAAGGAAGTTACTCTCTGTGACATCTTGGATGTAAACACTTGGATTGTGATA 169
1532 GAATAACCCATTGAATTTCTGCTGCGAGGGTGTAGAAATTTACTTTTGGGTATA 1591
168 GAATAACCCATTGAATTTCTGCTGCGAGGGTGTAGAAATTTACTTTTGGGTATA 109
1592 TTCTTATATATATATGATACATCGCTGCTCTGAAATTTAGTTATTTTGTGTTTAAATAA 1651
108 TTCTTATATATATATGATACATCGCTGCTCTGAAATTTAGTTATTTTGTGTTTAAATAA 49
1652 AGACTAACACAACTTAATGATTAAAGTGA 1682
48 AGACTAACACAACTTAATGATTAAAGTGA 18

RESULT 8
CB114674
LOCUS      K-EST0158382 L10ChOCK0 Homo sapiens cDNA clone L10ChOCK0-1-D02 5',
            mRNA sequence.
DEFINITION K-EST0158382 L10ChOCK0 Homo sapiens cDNA clone L10ChOCK0-1-D02 5',
            mRNA sequence.
ACCESSION  CB114674
VERSION     CB114674.1 GI:27940481
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
            21C Frontier Korean EST Project 2001
            Unpublished
            Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoseun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 1 row: D column: 02
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source           1..643
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L10ChOCK0-1-D02"
                     /sex="M"

```

BASE COUNT		214	a	123	c	126	g	180	t
ORIGIN									
Query Match		31.9%; Score 636; DB 14; Length 643;							
Best Local Similarity		100.0%; Pred. No. 2.4e-210;							
Matches		636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
Qy	894	GAC	CAG	AAC	TG	AC	AT	CT	CAG
Db	8	GAC	CAG	AAC	TG	AC	AT	CT	CAG
Qy	954	TG	TT	A	A	A	A	A	A
Db	68	TG	TT	A	A	A	A	A	A
Qy	1014	TT	AT	G	A	G	G	T	AT
Db	128	TT	AT	G	A	G	G	T	AT
Qy	1074	TT	TT	C	A	G	T	T	C
Db	188	TT	TT	C	A	G	T	T	C
Qy	1134	AC	AC	CT	T	G	G	A	A
Db	248	AC	AC	CT	T	G	G	A	A
Qy	1194	AG	C	A	C	C	T	T	C
Db	308	AG	C	A	C	C	T	T	C
Qy	1254	A	A	C	T	G	C	T	C
Db	368	A	A	C	T	G	C	T	C
Qy	1314	A	T	T	A	C	A	T	T
Db	428	A	T	T	A	C	A	T	T
Qy	1374	G	A	C	T	T	G	T	T
Db	488	G	A	C	T	T	G	T	T
Qy	1434	T	G	G	A	C	T	T	T
Db	548	T	G	G	A	C	T	T	T
Qy	1494	C	T	G	A	C	T	T	T
Db	608	C	T	G	A	C	T	T	T

RESULT 9  
EG256631  
LOCUS  
DEFINITION  
          BG256631          1105 bp      linear      EST 13-FEB-2001  
          602370809F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:447858 5',  
          mRNA sequence.  
ACCESSION  
VERSION      BG256631  
KEYWORDS      BG256631.1 GI:12766447  
SOURCE      EST.  
          Homo sapiens (human)  
ORGANISM     Homo sapiens  
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
1 (bases 1 to 1105)	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabbs@mail.nih.gov">cgabbs@mail.nih.gov</a> Tissue Procurement: ATCC		CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. CDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM10310 row: a column: 03 High quality sequence stop: 645. Location/Qualifiers 1. .1105 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4478858" /issue_type="embryonal carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC 92" /notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC library." 380 a 235 c 208 g 282 t
BASE COUNT				

	Query Match	31.1%;	Score 620;	DB 10;	Length 1105;
	Best Local Similarity	100.0%;	Prod. No. 6e-205;		
	Matches 620;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	764	CCAGAGTAAATCTTGGGAAAATTTAAAAATCTGTTCTCTGGAAAACTAAATGAAGAGAAA	823		
Db	1	CCAGATGTAATCTTGGGAAAATTTAAAAATCTGTTCTCTGGAAAACTAAATGAAGAGAAA	60		
Qy	824	GGTCAGCANTTGGAAACCTTTTATCATGAATTTTCATTAATTTTGTGAGTCTCCAAAGCCT	883		
Db	61	GGTCAGCANTTGGAAACCTTTTATCATGAATTTTCATTAATTTTGTGAGTCTCCAAAGCCT	120		
Qy	884	AAACCAAGTAGACCAAGACTGACCAATCTCAGCGCCTACTTTCAGAAAAACAACAAGAAGCTT	943		
Db	121	AAACCAAGTAGACCAAGACTGACCAATCTCAGCGCCTACTTTCAGAAAAACAACAAGAAGCTT	180		
Qy	944	TTCAATGATCTGTTTAAAAAATAATGCAAAACCGTGTCTGAAAAATACAGAGAGAAAGCAAAAT	1003		
Db	181	TTCAATGATCTGTTTAAAAAATAATGCAAAACCGTGTCTGAAAAATACAGAGAGAAAGCAAAAT	240		
Qy	1004	CAGAAATATTTTATGGAGGTGATGACTGTAGAAAGAGTCTATGATTAACCTGATGTATGTA	1063		
Db	241	CAGAAATATTTTATGGAGGTGATGACTGTAGAAAGAGTCTATGATTAACCTGATGTATGTA	300		
Qy	1064	GGACGGGTAGTTTCCAGGTTCTCGACTGCTCTCATCATCTCTTAATGGGAACCTCGAATC	1123		
Db	301	GGACGGGTAGTTTCCAGGTTCTCGACTGCTCTCATCATCTCTTAATGGGAACCTCGAATC	360		
Qy	1124	CTCTTTAAAAACACCCTCGAAATGTATCTAGTATTACTTTCAGTGTAAACTAGAACAG	1183		
Db	361	CTCTTTAAAAACACCCTCGAAATGTATCTAGTATTACTTTCAGTGTAAACTAGAACAG	420		
Qy	1184	CTATTTTCAGAGACACCGTTTGGTCTCACTCATTAACACTTCTCAGAGATGCTATATCTGT	1243		
Db	421	CTATTTTCAGAGACACCGTTTGGTCTCACTCATTAACACTTCTCAGAGATGCTATATCTGT	480		
Qy	1244	GAAAACTCTGAACTCTCGCTCTCTCCAAAGTAAGCAAAAAAGGAGCAAAACAGACTTTTGA	1303		
Db	481	GAAAACTCTGAACTCTCGCTCTCTCCAAAGTAAGCAAAAAAGGAGCAAAACAGACTTTTGA	540		
Qy	1304	GAAATGATGAATTAACATTCAGATCTGTGTAGTCAAGTGTATTTGGTGAAGAAAACCAAGTAT	1363		



```

541 GAAATGATGATTAATACATTCAGATCTGTAGTCAAGTGTATTGGTGAAGAAACCAAGTAT 600
1364 GAAAGCATCAGACTTCTGTT 1383
601 GAAAGCATCAGACTTCTGTT 620

CA431124 802 bp mRNA linear EST 07-NOV-2002
UI-H-FLI-bge-f-14-0-UI-si NCI CGAP FLI Homo sapiens cDNA clone
UI-H-FLI-bge-f-14-0-UI 3', mRNA sequence.
CA431124
CA431124.1 GI:24793850
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 802)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-70, >AT richLow_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..802
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FLI-bge-f-14-0-UI"
                     /tissue_type="Cell lines"
                     /dev_stage="Adult"
                     /lab_hosts="DH10B (Life Technologies)"
                     /clone_lib="NCI CGAP FLI"
                     /note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia
I; NCI CGAP FLI is a normalized cDNA library derived from
a pool of mRNA obtained from 4 cell lines from grade III
chondrosarcoma tissues. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p773-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GAGTCGGTG. The cell
lines were provided by Dr. James Martin from the
University of Iowa.
TAG_LIB=UI-H-FLI
TAG_TISSUE=Human Chondrosarcoma Grade 3 cell line mix
TAG_SEQ=GAGTCGGTG"

BASE COUNT          238 a 151 c 136 g 275 t 2 others
ORIGIN

Query Match          31.1%; Score 619; DB 14; Length 802;
Best Local Similarity 99.7%; Pred. No. 1.6e-204;
Matches 719; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 11
BUS79763/c
LOCUS
DEFINITION
BUS79763
ACCESSION
BUS79763
VERSION
BUS79763.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 629)
REFERENCE
1
AUTHORS
Mellon, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blais, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium

958 TAAAAATATGCAACCGTCTGAAATACAGAGAGAAAGCAAAATCAGATTTATTTTAT 1017
735 TAAAAATATGCAACCGTCTGAAATACAGAGAGAAAGCAAAATCAGATTTATTTAT 676
1018 GGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGTAGTTT 1077
675 GGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGTAGTTT 616
1078 CCAGGTTCTGACTGGCTTCATCTCTTAATGGGAACCTCGAATCTCTTTAAAAACAC 1137
615 CCAGGTTCTGACTGGCTTCATCTCTTAATGGGAACCTCGAATCTCTTTAAAAACAC 556
1138 CCTGGAATCTATACCTGATTACTCTTCAGTCTAACTAGAACAGCTATTTCAGAGACA 1197
555 CCTGGAATCTATACCTGATTACTCTTCAGTCTAACTAGAACAGCTATTTCAGAGACA 496
1198 CCGTTTGGTCTCACTCATACACTTCTCAGAGATGCTATATTCTGTGAAACACTGAACC 1257
495 CCGTTTGGTCTCACTCATACACTTCTCAGAGATGCTATATTCTGTGAAACACTGAACC 436
1258 TCGTCTCTCCAGATGAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAGATGATGAATTA 1317
435 TCGTCTCTCCAGATGAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAGATGATGAATTA 376
1318 CATTCCAGATCTGTAGTCAAGTGTATTGGTGAAGAACCAAGTATGAAAGCATCAGACT 1377
375 CATTCCAGATCTGTAGTCAAGTGTATTGGTGAAGAACCAAGTATGAAAGCATCAGACT 316
1378 TCTGTTTGGTCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTTTATTTGGA 1437
315 TCTGTTTGGTCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTTTATTTGGA 256
1438 CATTGTCATACAGAACTGTTTCCAGAGCTCAATAAGGTCACAAAGAAAGTACCTCTGT 1497
255 CATTGTCATACAGAACTGTTTCCAGAGCTCAATAAGGTCACAAAGAAAGTACCTCTGT 196
1498 GACATCTGGATGTAAACACTTGGATTTGGTATAGATACCCATTTGAATTTCTGCTGT 1557
195 GACATCTGGATGTAAACACTTGGATTTGGTATAGATACCCATTTGAATTTCTGCTGT 136
1558 GCGAGGTGGTGTAGAAATTTACTTTTTGGTATATTCTTATATATATATATGATCATCGCT 1617
135 GCGAGGTGGTGTAGAAATTTACTTTTTGGTATATTCTTATATATATATATGATCATCGCT 76
1618 GTCTGAAATTTAGTATTTTGTGTTTTTAATAGACTAACACAACTTAATGATTTAA 1677
75 GTCTGAAATTTAGTATTTTGTGTTTTTAATAGACTAACACAACTTAATGATTTAA 16
1678 A 1678
15 A 15

BUS79763 629 bp mRNA linear EST 17-SEP-2002
im90e05.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:612216
3' similar to TR:Q9Y5W7 Q9Y5W7 SORTING NEXIN 14 ; mRNA sequence.

BUS79763
ACCESSION
BUS79763
VERSION
BUS79763.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 629)
REFERENCE
1
AUTHORS
Mellon, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blais, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium

```

JOURNAL  
COMMENT

Unpublished  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@bichp.harvard.edu  
Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
University Genome Sequencing Center For information on obtaining a  
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 456.

## FEATURES

Location/Qualifiers  
source  
1..629  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:612216"  
/issue\_type="insulinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="Human insulinoma"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system  
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue following the Washington  
University protocol  
(http://genome.wustl.edu/est/lambda\_protocol.shtml).  
Please contact Hiroshi Inoue, MD/PhD for further  
information on this library (Metabolism Division, Permutt  
Laboratory, Washington University School of Medicine, Box  
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this  
is a Washington University Pancreas EST project library."

BASE COUNT 198 a 113 c 107 g 211 t  
ORIGIN

Query Match 30.8%; Score 613; DB 13; Length 629;  
Best Local Similarity 100.0%; Pred. No. 2.3e-202;  
Matches 613; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1070 GTAGTTTTCAGGTTCTGACTGGCTTCATCATCTTTAATGGAACTCGAATCTCTTTT 1129  
DB 629 GTAGTTTTCAGGTTCTGACTGGCTTCATCATCTTTAATGGAACTCGAATCTCTTTT 570  
QY 1130 AAAAACCCTCGAATGTATCTAGTATTACTATCTTCAGTGTAACAGCAGCTATTT 1189  
DB 569 AAAAACCCTCGAATGTATCTAGTATTACTATCTTCAGTGTAACAGCAGCTATTT 510  
QY 1190 CAGGAGCACCCTTTGGTCTCCTCATAACACTTCTCAGAGATGCTATTTCTGTGAAC 1249  
DB 509 CAGGAGCACCCTTTGGTCTCCTCATAACACTTCTCAGAGATGCTATTTCTGTGAAC 450  
QY 1250 ACTGAACCTCGCTCTCTCCAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATG 1309  
DB 449 ACTGAACCTCGCTCTCTCCAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATG 390  
QY 1310 ATGAATTACATTCAGATCTGTTAGTCAAGTGATTTGGTGAAGAAACCAAGTATGAAGC 1369  
DB 389 ATGAATTACATTCAGATCTGTTAGTCAAGTGATTTGGTGAAGAAACCAAGTATGAAGC 330  
QY 1370 ATCAGACTTCGTTGATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTT 1429  
DB 329 ATCAGACTTCGTTGATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTT 270  
QY 1430 TTATTGGACATTTGTGATACAGGAACCTGTTTCCAGAGCTCAATAAGGTACAAAAGGAAGTT 1489  
DB 269 TTATTGGACATTTGTGATACAGGAACCTGTTTCCAGAGCTCAATAAGGTACAAAAGGAAGTT 210  
QY 1490 ACCTCTGTGACATCTTGGATGTAACACACTTGGATTTGGTATAGATAACCCATTGAATTT 1549  
DB 209 ACCTCTGTGACATCTTGGATGTAACACACTTGGATTTGGTATAGATAACCCATTGAATTT 150

QY 1550 TCTGCTGCTGCGAGGCTGTAGAAATTTACTTTTTTGGGTATATTTCTATATATATATATGT 1609  
DB 149 TCTGCTGCTGCGAGGCTGTAGAAATTTACTTTTTTGGGTATATTTCTATATATATATATGT 90  
QY 1610 ACATCGCTGCTGCTGAATTTTAGTATTATTTTGTATTAATAAGACTTAACACAACTTAA 1669  
DB 89 ACATCGCTGCTGCTGAATTTTAGTATTATTTTGTATTAATAAGACTTAACACAACTTAA 30  
QY 1670 TGATTAAGTGA 1682  
DB 29 TGATTAAGTGA 17

## RESULT 12

BM829345 608 bp mRNA linear EST 06-MAR-2002  
LOCUS K-EST0102319 S9SNU601 Homo sapiens cDNA clone S9SNU601-45-A08 5',  
DEFINITION mRNA sequence.  
ACCESSION BM829345  
VERSION BM829345.1 GI:19185754  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 608)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

## JOURNAL

COMMENT Unpublished  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoseun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 45' row: A column: 08  
High quality sequence stop: 608.  
Location/Qualifiers

## FEATURES

source

1..608  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S9SNU601-45-A08"  
/sex="M"  
/tissue\_type="Ascites"  
/cell\_type="Epithelial"  
/cell\_line="SNU-601"  
/lab\_host="Top10P"  
/clone\_lib="S9SNU601"  
/note="Organ: Stomach; Vector: pME18-FL3, Site 1: XhoI;  
Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DraIII- digested pME18-FL3 vector. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10P' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

BASE COUNT 203 a 100 c 130 g 175 t

## ORIGIN

Query Match 30.5%; Score 608; DB 12; Length 608;  
Best Local Similarity 100.0%; Pred. No. 1.3e-200;

Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 TAGCAAAATTAAGAGTATTCAAAAGTACCAATGAGGAGGAGTATGTCCTAAATTA 252

Db 1 TAGCAAAATTAAGAGTATTCAAAAGTACCAATGAGGAGGAGTATGTCCTAAATTA 60

QY 253 TGTGTAGCTGAGGTGAGATGATTTTATGGAAGGTATTTGTTGTAATGGAAGATGA 312

Db 61 TGTGTAGCTGAGGTGAGATGATTTTATGGAAGGTATTTGTTGTAATGGAAGATGA 120

QY 313 TTCTCCAGTGGAGCTGTGAGCACACTAATCTCCCGAAACCTTGCTGCAATGGAAT 372

Db 121 TTCTCCAGTGGAGCTGTGAGCACACTAATCTCCCGAAACCTTGCTGCAATGGAAT 180

QY 373 TAGCATTCATATGTAGACTTTTTTGGAGTCCCTCTCTGAAAGGAGGAGAAAAGA 432

Db 181 TAGCATTCATATGTAGACTTTTTTGGAGTCCCTCTCTGAAAGGAGGAGAAAAGA 240

QY 433 AAGNATTCCTGTTTGTATGATGTTGAAGAAATGATAGAGAGCAGTTGGACAGA 492

Db 241 AAGNATTCCTGTTTGTATGATGTTGAAGAAATGATAGAGAGCAGTTGGACAGA 300

QY 493 GCCTGAACATTTGCTCTATAGAGATATCTTGAATCTATGATCTTGAATCAAACT 552

Db 301 GCCTGAACATTTGCTCTATAGAGATATCTTGAATCTATGATCTTGAATCAAACT 360

QY 553 AACAGAAATTCATGGTCATTTCTGATGCCAGCTTCCTTTTAAGAGATCATTTGGCCC 612

Db 361 AACAGAAATTCATGGTCATTTCTGATGCCAGCTTCCTTTTAAGAGATCATTTGGCCC 420

QY 613 CAAAATATGAAATCTTAAGTCAAGAGGAGAGTTCACAGATATCTACAGAACT 572

Db 421 CAAAATATGAAATCTTAAGTCAAGAGGAGAGTTCACAGATATCTACAGAACT 480

QY 673 TCTCAGCATCCAGACTGATGATAGTCACTTCTGGCAGCTTTCTTTCCCTAATGG 732

Db 481 TCTCAGCATCCAGACTGATGATAGTCACTTCTGGCAGCTTTCTTTCCCTAATGG 540

QY 733 TGGGGAACACAAATTTCTGATAGATCTACAGATGTAATCTTGGGAAATTTATAA 792

Db 541 TGGGGAACACAAATTTCTGATAGATCTACAGATGTAATCTTGGGAAATTTATAA 600

QY 793 ATCTGTTTC 800

Db 601 ATCTGTTTC 608

RESULT 13

BM699448

LOCUS

DEFINITION

UI-E-DX1-agy-g-17-0-UI-r1 UI-E-DX1 Homo sapiens cDNA clone

UI-E-DX1-agy-g-17-0-UI 5', mRNA sequence.

ACCESSION

BM699448

VERSION

BM699448.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

MEDLINE

9704477

8889548

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEERF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Cloning Distribution: Dr. M. Bento Soares, University of Iowa

Cloning Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1. 606

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-E-DX1-agy-g-17-0-UI"

/tissue\_type="fetal eyes"

/dev\_stage="fetal"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-E-DX1"

/notes="Organ: eye; Vector: p77T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DX1 is a normalized cDNA library containing the following tissue(s): fetal eyes. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p77T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCAAGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 204 a 114 c 118 g 170 t

ORIGIN

Query Match 30.4%; Score 606; DB 12; Length 606;

Best Local Similarity 100.0%; Pred. No. 6.3e-200; Indels 0; Gaps 0;

Matches 606; Conservative 0; Mismatches 0;

QY 913 CAGCCCTACTTCAGAAAACAAAGAGCTTTTCAATGATCTGTTTAAAAATATGCAAA 972

Db 1 CAGCCCTACTTCAGAAAACAAAGAGCTTTTCAATGATCTGTTTAAAAATATGCAAA 60

QY 973 CCGTCTGAAAATACAGAGAGAAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGT 1032

Db 61 CCGTCTGAAAATACAGAGAGAAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGT 120

QY 1033 AGAAGAGCTATGATTTACCTGATGATGATGAGGAGGAGTATTTCCAGGTTCCTGACTG 1092

Db 121 AGAAGAGCTATGATTTACCTGATGATGATGAGGAGGAGTATTTCCAGGTTCCTGACTG 180

QY 1093 GCTTCATCATCTCTTAATGGGAACCTCGAATCTCTTTAAAAACACCCCTGGAAATGTATAC 1152

Db 181 GCTTCATCATCTCTTAATGGGAACCTCGAATCTCTTTAAAAACACCCCTGGAAATGTATAC 240

QY 1153 TGATTTACTATCTTTCAGTGTAAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCACT 1212

Db 241 TGATTTACTATCTTTCAGTGTAAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCACT 300

QY 1213 CATACACTTCTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCGCTCTCTCCAGA 1272

Db 301 CATACACTTCTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCGCTCTCTCCAGA 360

QY 1273 TAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAATGATGATTAATTCAGATCTCTT 1332

Db 361 TAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAATGATGATTAATTCAGATCTCTT 420

QY 1333 AGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAGATCAGACTTCTGTTTGTGGCTT 1392

Db 421 AGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAGATCAGACTTCTGTTTGTGGCTT 480

QY 1393 ACAGCAACCAGTACTCAACAGCAGCTGACTTATGTTTTTATTTGGACATTTGTGATACAGA 1452

Db	481	ACGACCAACAGTACTCAGACAGCTGACTTATGTTTATTTGGACATTTGGATACAGA	540
Qy	1453	ACTGTTTCCAGAGCTCAATAAGGTACAAAGGAAGTACCTCTGTGACATCTTGGATGTA	1512
Db	541	ACTGTTTCCAGAGCTCAATAAGGTACAAAGGAAGTACCTCTGTGACATCTTGGATGTA	600
Qy	1513	AACACT 1518	
Db	601	AACACT 606	
RESULT 14			
LOCUS	CB130920	604 bp mRNA linear EST 29-JAN-2003	
DEFINITION	K-EST0180911 L12JSHCO Homo sapiens cDNA clone L12JSHCO-5-A04 5', mRNA sequence.		
ACCESSION	CB130920		
VERSION	CB130920.1	GI:28095369	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 604) Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.S., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.		
TITLE	21C Frontier Korean EST Project 2001		
JOURNAL	Unpublished		
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongseung@mail.kribb.re.kr Plate: 5 row: A column: 04 High quality sequence stop: 604.		
FEATURES	Location/Qualifiers		
source	1..604		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="L12JSHCO-5-A04"		
	/sex="M"		
	/cell_line="J-SHC"		
	/lab_host="L12JSHCO"		
	/note="Organ: Liver; Vector: pcNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dr-selected mRNA by priming with dr-tailed vector. The dr-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transfection of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."		
BASE COUNT	205 a	100 c	137 g
ORIGIN	162 t		
Query Match	30.1%	Score 599;	DB 14; Length 604;
Best Local Similarity	100.0%;	Pred. No. 1.7e-197;	
Matches 599;	Conservative	0; Mismatches	0; Indels
		0; Gaps	0;
Qy	145	CAGGACACACAGAAAGGGGAGATCATTTTGGATCAGCAGATAGGTAGCAAAATTA	204

POLYA=Yes. Location/Qualifiers  
1. .594  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DWI-aha-m-13-0-UI"  
/tissue\_type="lens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DWI"  
/note="Organ: eye; Vector: pMT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DWI is a normalized cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pMT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
TAG LIB=UI-E-DWI  
TAG TISSUE=human lens  
TAG\_SEQ=CGATTAGCGA" 207 t

BASE COUNT 184 a 106 c 97 g 207 t  
ORIGIN

Query Match 29.0%; Score 577; DB 13; Length 594;  
Best Local Similarity 100.0%; Pred. No. 7.4e-190;  
Matches 577; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1106 TTAATGGAACTCGAATCTCTTTAAACACCCCTGGAAATGATGATTACTATCTT 1165  
Db 594 TTAATGGAACTCGAATCTCTTTAAACACCCCTGGAAATGATGATTACTATCTT 535  
Qy 1166 CAGGTAACTAGAACAGCTATTTCAGGACGCGTTGGTCTCCTCATACACTTC 1225  
Db 534 CAGGTAACTAGAACAGCTATTTCAGGACGCGTTGGTCTCCTCATACACTTC 475  
Qy 1226 AGAGATGCTATATCTGTGAAACACCTGAACCTCGCTCTCCAGATAAGCAAAAGGA 1285  
Db 474 AGAGATGCTATATCTGTGAAACACCTGAACCTCGCTCTCCAGATAAGCAAAAGGA 415  
Qy 1286 GCAAAACAGACTTTTGAAGAAATGATGAATTACATTCCAGATCTGTAGTCAAGTGATT 1345  
Db 414 GCAAAACAGACTTTTGAAGAAATGATGAATTACATTCCAGATCTGTAGTCAAGTGATT 355  
Qy 1346 GGTGAAGAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACGACCAAGTA 1405  
Db 354 GGTGAAGAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACGACCAAGTA 295  
Qy 1406 CTCACAGCAGCTGACTTATGTTTATTCGACATTTGTGATACAGGACTGTTTCCAGAG 1465  
Db 294 CTCACAGCAGCTGACTTATGTTTATTCGACATTTGTGATACAGGACTGTTTCCAGAG 235  
Qy 1466 CTCATTAAGGTACAAAAGGAGTTTACCTCTGTGACATCTTGGATGTAACACCTGGATT 1525  
Db 234 CTCATTAAGGTACAAAAGGAGTTTACCTCTGTGACATCTTGGATGTAACACCTGGATT 175  
Qy 1526 GGTATAGATAACCCATGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTTTTG 1585  
Db 174 GGTATAGATAACCCATGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTTTTG 115  
Qy 1586 GGTATATTCCTATATATATATATGATACATCGCTGCTGAAATTTTACTTTTTTCTTTT 1645  
Db 114 GGTATATTCCTATATATATATATGATACATCGCTGCTGAAATTTTACTTTTTTCTTTT 55  
Qy 1646 TAATAAGACTAACACAACTTAATGATTAAGTGA 1682

Db 54 TAATAAGACTTAACACAACTTAATGATTAAGTGA 18

Search completed: January 31, 2004, 19:32:41  
Job time : 4308 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 13:56:38 ; Search time 548 Seconds  
(without alignments)  
9812.556 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1-gratgaactccaaagt.....tttaataataaaaaaaaa 1992

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapexb\_60.0

Searched: 2552756 seqs 1349719037 residues

Word size: 0  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1992	100.0	1992	22 AAC86397	SNEXN DNA #1. Hom
C 2	1838	92.3	2176	22 AA160815	Human polynucleoti
C 3	1838	92.3	2176	22 AA160816	Human polynucleoti
4	1532	76.9	1716	22 AA159029	Human polynucleoti
5	1296	65.1	3145	24 ABQ79518	cDNA encoding a pr
6	918	46.1	1551	22 AA159030	Human polynucleoti
7	543	27.3	779	20 AA217379	Human gene express
C 8	385	19.3	451	24 ABL63556	Breast cancer rela

C 9	357	17.9	358	21 AAC02755	Human secreted pro
10	78	3.9	580	24 ABL62287	Human cancer relat
11	76	3.8	311	22 AAH69552	Human cervical can
12	76	3.8	313	22 AAH70887	Human cervical can
13	74	3.7	299	22 AAH72445	Human cervical can
14	74	3.7	299	22 AAH73007	Human cervical can
15	60	3.0	60	24 ABL39970	Human spliced tran
16	59	3.0	305	25 ABL50426	Bovine EST associa
17	59	3.0	402	25 ABL50611	Bovine EST associa
C 18	23	1.2	39380	22 AAH74027	Human immune/haema
19	22	1.1	5807	24 ABL33155	Human immune syste
20	22	1.1	5815	24 ABL70585	Chemically treated
21	22	1.1	6009	24 ABL32829	Human immune syste
22	22	1.1	6070	24 ABL67129	Human angiogenesis
23	22	1.1	6070	24 ABL70371	Chemically treated
24	22	1.1	6070	24 ABL33678	Human immune syste
25	22	1.1	6070	24 ABL34578	Human metastasis a
C 26	22	1.1	6277	22 AAS46324	Tumour suppressor
27	22	1.1	7049	24 ABL54306	Chemically treated
28	22	1.1	7049	24 ABL32157	Human immune syste
29	22	1.1	7058	24 ABL40052	Human chemically p
30	22	1.1	7238	24 ABL32449	Human immune syste
31	22	1.1	7771	24 ABL33973	Human immune syste
C 32	22	1.1	8056	25 ABL210100	Haematopoietic cel
C 33	22	1.1	8056	25 ABL210246	Haematopoietic cel
34	22	1.1	8076	24 ABL39955	Human chemically p
35	22	1.1	8576	24 ABL34229	Human immune syste
36	22	1.1	10250	24 ABL80082	Human immune syste
37	22	1.1	10716	24 ABL33419	Human immune syste
38	22	1.1	13511	24 ABL32280	Human immune syste
C 39	22	1.1	14798	24 ABL33032	Chemically treated
40	22	1.1	19459	24 ABL70527	Chemically treated
41	22	1.1	19459	24 ABL31212	Signal transductio
42	22	1.1	33053	24 ABL67005	Human angiogenesis
43	22	1.1	34548	24 ABL70603	Chemically treated
44	22	1.1	40862	24 ABL34072	Human immune syste
45	22	1.1	73334	24 ABL92318	Chemically treated

#### ALIGNMENTS

#### RESULT 1

AAC86397  
ID AAC86397 standard; DNA; 1992 BP.  
AC AAC86397;  
XX  
DT 21-MAR-2001 (first entry)  
XX  
DE SNEXN DNA #1.  
XX  
KW SNEXN; human; sorting nexin; inflammation; asthma; allergy; AIDS;  
KW neurological disorder; gastrointestinal; smooth muscle cell;  
KW cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO2000073334-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 26-MAY-2000; 2000WO-US14831.  
XX  
PR 27-MAY-1999; 99US-0136740.  
PR 16-JUN-1999; 99US-0139566.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
PI Yue H, Tang YT, Azimzai Y;  
XX WPI; 2001-041141/05.  
XX



1801 GACTGATAAAGCTTGGTATAGCTTGAATAAATATATGCTAGTATGGAGAAACAG 1860  
 1861 GAATAAGATCTGATTTCTAGAGTAAATATATTTAGTAGATTTGGTTTCTTTT 1920  
 1861 GAATAAGATCTGATTTCTAGAGTAAATATATTTAGTAGATTTGGTTTCTTTT 1920  
 1921 ATTTTGTACATAGTAACTGCTATCTATATAAATAAGCATCTATATGAGTTTAAATAA 1980  
 1921 ATTTTGTACATAGTAACTGCTATCTATATAAATAAGCATCTATATGAGTTTAAATAA 1980  
 1981 TAAAAAATAA 1992  
 1981 TAAAAAATAA 1992

RESULT 2  
 AAI60815/c  
 ID AAI60815 standard; cDNA; 2176 BP.  
 XX  
 AC AAI60815;  
 XX  
 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 4804.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;  
 KW Chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 26-DEC-2000; 2000WO-US34263.  
 XX  
 XX 21-JAN-2000; 2000US-0488725.  
 XX 25-APR-2000; 2000US-0552317.  
 XX 09-JUL-2000; 2000US-0598042.  
 XX 19-JUL-2000; 2000US-0620312.  
 XX 03-AUG-2000; 2000US-0653450.  
 XX 14-SEP-2000; 2000US-0662191.  
 XX 19-OCT-2000; 2000US-0693036.  
 XX 29-NOV-2000; 2000US-0727344.  
 (HVS8-) HVS8Q INC.  
 XX  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 XX WPI: 2001-442253/47.  
 XX P-PSDB; AAM1659.  
 XX  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 XX such as central nervous system injuries -  
 XX  
 XX Claim 1; SEQ ID NO 4804; 10078pp; English.  
 XX  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
 XX in gene therapy. A composition containing a polypeptide or polynucleotide  
 XX of the invention may be used to treat diseases of the peripheral nervous  
 XX system, such as peripheral nervous injuries, peripheral neuropathy and  
 XX localised neuropathies and central nervous system diseases, such as  
 XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 XX utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
 Query Match 92.3%; Score 1838; DB 22; Length 2176;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 147 GGACACACAGAAAGGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATTAAG 206  
 DB 2018 GGACACACAGAAAGGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATTAAG 1959  
 QY 207 GAGTATTCAAAGTACCAATCGAGGAGCTATGTTGCTAATTAATGTTAGCTGAAG 266  
 DB 1958 GAGTATTCAAAGTACCAATCGAGGAGCTATGTTGCTAATTAATGTTAGCTGAAG 1899  
 QY 267 GTGAAGATGATTTTATTGAAGAGGTATTGTTGTAATGGAAGATGATTTCTCAGTGAGG 326  
 DB 1898 GTGAAGATGATTTTATTGAAGAGGTATTGTTGTAATGGAAGATGATTTCTCAGTGAGG 1839  
 QY 327 CTGTGACACACCTAATATCTCCCGAAACCTTCTCTCATGGAATATAGCATTCATATG 386  
 DB 1838 CTGTGACACACCTAATATCTCCCGAAACCTTCTCTCATGGAATATAGCATTCATATG 1779  
 QY 387 TAGCATTTTGTAGGATCCCTCTCTGAAAGGAGAGAGAAAAGAAAGAAATTCCTGTGT 446  
 DB 1778 TAGCATTTTGTAGGATCCCTCTCTGAAAGGAGAGAGAAAAGAAAGAAATTCCTGTGT 1719  
 QY 447 TTTGTTATTGATTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGAAACATTTGT 506  
 DB 1718 TTTGTTATTGATTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGAAACATTTGT 1659  
 QY 507 CTGTCTATAGAGATATCTTGAATTTCTATGTAATTGAATCAAAACCTAAACAGATTCATG 566  
 DB 1658 CTGTCTATAGAGATATCTTGAATTTCTATGTAATTGAATCAAAACCTAAACAGATTCATG 1599  
 QY 567 GTGCATTTCTGTATGCCACCTTCTTCTTAAGAGGATCATTTGCCCCCAAAATTTGAAT 626  
 DB 1598 GTGCATTTCTGTATGCCACCTTCTTCTTAAGAGGATCATTTGCCCCCAAAATTTGAAT 1539  
 QY 627 TCTTAAAGTCAAGAGAGGAGAGGTTCCAGAAATATCTACAGAACTTCTCAGCATCCAG 686  
 DB 1538 TCTTAAAGTCAAGAGAGGAGAGGTTCCAGAAATATCTACAGAACTTCTCAGCATCCAG 1479  
 QY 687 AACTGAGTAAATAGTCAACTTTCTGCGACACTTTCTTCCCTTAATGTTGGGAAACACAAT 746  
 DB 1478 AACTGAGTAAATAGTCAACTTTCTGCGACACTTTCTTCCCTTAATGTTGGGAAACACAAT 1419  
 QY 747 TTTCTGTATAGATACTACAGATGTAATCTTGGGAAATTAATAATCTGTTCTCTGGAA 806  
 DB 1418 TTTCTGTATAGATACTACAGATGTAATCTTGGGAAATTAATAATCTGTTCTCTGGAA 1359  
 QY 807 AACTAATGAAAGAGAAAGGTCAGCATTTGGAACTTTATCATGAATTTCTCAATTAATCTT 866  
 DB 1358 AACTAATGAAAGAGAAAGGTCAGCATTTGGAACTTTATCATGAATTTCTCAATTAATCTT 1299  
 QY 867 GTGAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTTCAG 926  
 DB 1298 GTGAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTTCAG 1239  
 QY 927 AAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAAATA 986  
 DB 1238 AAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAAATA 1179  
 QY 987 CAGAGAGAAACCAAAATCAGAAATTTATTTATGAGGATGATCTGTAGAGAGTCTATG 1046  
 DB 1178 CAGAGAGAAACCAAAATCAGAAATTTATTTATGAGGATGATCTGTAGAGAGTCTATG 1119





1898	GTGAAAGATGATTTTATTGTAAGAAGGTATTGTGTANTGAAGATGATTTCTCCAGTGGAGG	1833
327	CTGTGAGCACACTAATACTCCCGAAAACCTTGCTGCATGGAATAATAGCATTTCCATATG	386
1838	CTGTGAGCACACTAATACTCCCGAAAACCTTGCTGCATGGAATAATAGCATTTCCATATG	1779
387	TAGACATTTTTTGGAGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTCT	446
1778	TAGACATTTTTTGGAGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTCT	1719
447	TTTGTATTGATGTTGAAAGAAATGATAGAAAGACAGTTGGACACAGAGCCTGAAACATTCGT	506
1718	TTTGTATTGATGTTGAAAGAAATGATAGAAAGACAGTTGGACACAGAGCCTGAAACATTCGT	1659
507	CTGCTATAGACATATCTTTGAATCTATGATCTTGAATCAAAACTAAACAGAAATTCATG	566
1658	CTGCTATAGACATATCTTTGAATCTATGATCTTGAATCAAAACTAAACAGAAATTCATG	1599
567	GTGCAATTCCTGATGCCAGCTTCCTCTTAAGAGGATCATTTGGCCCCCAAAAATTTATGAAT	626
1598	GTGCAATTCCTGATGCCAGCTTCCTCTTAAGAGGATCATTTGGCCCCCAAAAATTTATGAAT	1539
627	TCCTTAAGTCAAAAGAGGGAAGGTTCCAGAAATATCTACAGAAACTTTCTGCAGCATCCAG	686
1538	TCCTTAAGTCAAAAGAGGGAAGGTTCCAGAAATATCTACAGAAACTTTCTGCAGCATCCAG	1479
687	AACTGAGTAATAGTCAACTCTCGCGACATTTCTTTCCCTTAATGTGGGGAACAACAAT	746
1478	AACTGAGTAATAGTCAACTCTCGCGACATTTCTTTCCCTTAATGTGGGGAACAACAAT	1419
747	TTCTTTGATAGATATCTACAGATGTAAATCTTGGGAAAAATTTATAAATCTGTTCCTGGAA	806
1418	TTCTTTGATAGATATCTACAGATGTAAATCTTGGGAAAAATTTATAAATCTGTTCCTGGAA	1359
807	AACATAATGAAGAGAAAGGTTCAGCATTTGGAACCTTTTATCATGAAATTTTCATTAATTCCT	866
1358	AACATAATGAAGAGAAAGGTTCAGCATTTGGAACCTTTTATCATGAAATTTTCATTAATTCCT	1299
867	GTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCAATCTTCAGGCCCTACTTCAG	926
1298	GTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCAATCTTCAGGCCCTACTTCAG	1239
927	AAAAACAACAAGCCTTTCAATGATCTGTTTTAAAAATAATATGCAAAACCGTGTGAAAAATA	986
1238	AAAAACAACAAGCCTTTCAATGATCTGTTTTAAAAATAATATGCAAAACCGTGTGAAAAATA	1179
987	CAGAGAGAAAGCAAAATCAGAAATTTTATGAGAGGTGATGCTGTAGAGAGGAGTCTATG	1046
1178	CAGAGAGAAAGCAAAATCAGAAATTTTATGAGAGGTGATGCTGTAGAGAGGAGTCTATG	1119
1047	ATTACCTGATGATGTAGGACGGGTAGTTTTTCCAGGTTCTCTGACTGGCCTTCATCATCTCT	1106
1118	ATTACCTGATGATGTAGGACGGGTAGTTTTTCCAGGTTCTCTGACTGGCCTTCATCATCTCT	1059
1107	TAATGGGAACCTCGAATCCTCTTTAAAAACAACCTCGGAAAATGATATCTATCTATCTTC	1166
1058	TAATGGGAACCTCGAATCCTCTTTAAAAACAACCTCGGAAAATGATATCTATCTATCTTC	999
1167	AGTGTAACCTAGAACAGCTATTTACAGGAGCACCGTTTGGTCTCACTCATTAACACTTCTCA	1226
998	AGTGTAACCTAGAACAGCTATTTACAGGAGCACCGTTTGGTCTCACTCATTAACACTTCTCA	939
1227	GAGATGCTATATCTGTGAAAACACTGAACTCGCTCTCTCCAAAGATAAGCAAAAAGGAG	1286
938	GAGATGCTATATCTGTGAAAACACTGAACTCGCTCTCTCCAAAGATAAGCAAAAAGGAG	879
1287	CAAAACAGACTTTTGAAGAAATGATGAATTAACAATCCAGATCTCTTTAGTCAAGTGATTTG	1346
878	CAAAACAGACTTTTGAAGAAATGATGAATTAACAATCCAGATCTCTTTAGTCAAGTGATTTG	819
1347	GTGAAGAAAACCAAGTATGAAAGCATCAGACTTTCTGTTTGGATGGCTTACAGCAACAGTAC	1406
818	GTGAAGAAAACCAAGTATGAAAGCATCAGACTTTCTGTTTGGATGGCTTACAGCAACAGTAC	759

#### RESULT 4

AAI59029

AA135029  
ID AAT59029 standard: cDNA: 1716 BP.

AC AAT59029:

22-OCT-2001 (first entry)

Human polynucleotide SEO ID NO 1232.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
anoythropic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukemia; ss.

XX Homo sapiens.

XX  
PN  
W0200153312-A1

XX  
26-.III.-2001.XX  
PF 26-DEC-2000: 2000WQ-UIS34263.XX  
PP 21-JAN-2000: 2000US-0488725.

PR 21-JAN-2000; 2000US=0488723.  
PR 25-APR-2000; 2000US=0552317.

PR 23-APR-2000; 2000US-0332317.  
PR 09-JUL-2000; 2000US-0598042.

```
PR 03-JUL-2000; 2000US-062305Z
PR 19-JUL-2000; 2000US-062305Z
```

PR 13-JUL-2000; 2000US-0653450;  
PR 03-AUG-2000; 2000US-0653450;

14-SEP-2000: 2000US-0662191:

PR 19-OCT-2000; 2000US-0693036.  
PR 23-NOV-2000; 2000US-0727344.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAM39873.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX Claim 1; SEQ ID NO 1232; 10078pp; English.  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;  
  
Query Match 76.98; Score 1532; DB 22; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 147 GGAACACACAGAAAGGGAGATCAATTTGGAAATCAGACAGATAGGTAGCAAAATTAAG 206  
DB 177 GGAACACACAGAAAGGGAGATCAATTTGGAAATCAGACAGATAGGTAGCAAAATTAAG 236  
QY 207 GAGTATTTCAAAGTACACAAATGAGGAGCTATGTTGCCATTAATATGTTAGTGAAG 266  
DB 237 GAGTATTTCAAAGTACACAAATGAGGAGCTATGTTGCCATTAATATGTTAGTGAAG 296  
QY 267 GTGAAGATGATTTATTTGAAGAGAGTATTTGTTATGTAAGATGATTTCCAGTGGAG 326  
DB 297 GTGAAGATGATTTATTTGAAGAGAGTATTTGTTATGTAAGATGATTTCCAGTGGAG 356  
QY 327 CTGTGACACACCTAATCTCCCGAAACCTTCTGCTGATCGAAATTAGCATTTCCATATG 386  
DB 357 CTGTGACACACCTAATCTCCCGAAACCTTCTGCTGATCGAAATTAGCATTTCCATATG 416  
QY 387 TAGACTTTTGTAGGATCCCTCTCTGAAAGAGAGAGAAAGAAAGAAATTCCTGTGT 446  
DB 417 TAGACTTTTGTAGGATCCCTCTCTGAAAGAGAGAGAAAGAAAGAAATTCCTGTGT 476  
QY 447 TTTGTATTTGATTTGAAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 506  
DB 477 TTTGTATTTGATTTGAAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 536  
QY 507 CTGTCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTAAGAGATTTTCATG 566  
DB 537 CTGTCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTAAGAGATTTTCATG 596  
QY 567 GTGCATTTCTGTAGTCCAGCTTCTCTTAAGAGAGATCATTTGGCCCAAAATTTATGAAT 626  
DB 597 GTGCATTTCTGTAGTCCAGCTTCTCTTAAGAGAGATCATTTGGCCCAAAATTTATGAAT 656  
QY 627 TCTTAAAGTCAAAGAGGAGAGATTTCCAAAGATATCTACAGAAATCTTCTGCAGCATCCAG 686

DB 657 TCTTAAAGTCAAAGAGGAGAGATTTCCAAAGATATCTACAGAAATCTTCTGCAGCATCCAG 716  
QY 687 AACTGAGTAAATAGTCAACTTCTGGCAGACTTTTCCCTAATGCTGGGAAACCAAT 746  
DB 717 AACTGAGTAAATAGTCAACTTCTGGCAGACTTTTCCCTAATGCTGGGAAACCAAT 776  
QY 747 TCTTTGATAGATATCTACAGATGTAATCTTGGGAAATTTAATAATCTGTTCTCTGGAA 806  
DB 777 TCTTTGATAGATATCTACAGATGTAATCTTGGGAAATTTAATAATCTGTTCTCTGGAA 836  
QY 807 AACTTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATTAATCTT 866  
DB 837 AACTTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATTAATCTT 896  
QY 867 GTGAGTCTCAAAGCCTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACTTCAG 926  
DB 897 GTGAGTCTCAAAGCCTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACTTCAG 956  
QY 927 AAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAAAATTAATCAAAACCGTCTGAAATA 986  
DB 957 AAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAAAATTAATCAAAACCGTCTGAAATA 1016  
QY 987 CAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGAGGTGATGACTGTAGAGAGAGTCTATG 1046  
DB 1017 CAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGAGGTGATGACTGTAGAGAGAGTCTATG 1076  
QY 1047 ATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1106  
DB 1077 ATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
QY 1107 TAATGGGAATCTCGAATCTCTTTAAAAACACCTTGGAAATGATGATGATGATGATGATGATG 1166  
DB 1137 TAATGGGAATCTCGAATCTCTTTAAAAACACCTTGGAAATGATGATGATGATGATGATGATG 1196  
QY 1167 AGTGTAACCTAGAACAGCTATTTTCCAGGAGCAGCTTTTGGTCTCTCACTCATCAACCTCTCA 1226  
DB 1197 AGTGTAACCTAGAACAGCTATTTTCCAGGAGCAGCTTTTGGTCTCTCACTCATCAACCTCTCA 1256  
QY 1227 GAGATGCTATATTTCTGTGAAACCACTGAACTGCTCTCTCCAGAGATGAGCAAAAGAGAG 1286  
DB 1257 GAGATGCTATATTTCTGTGAAACCACTGAACTGCTCTCTCCAGAGATGAGCAAAAGAGAG 1316  
QY 1287 CAAACACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1346  
DB 1317 CAAACACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1376  
QY 1347 GTGAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGGTCTGCTTACAGCAACCATGAC 1406  
DB 1377 GTGAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGGTCTGCTTACAGCAACCATGAC 1436  
QY 1407 TCAACCAAGCAGCTGACTTATGTTTATTTGGAATTTGATGATGATGATGATGATGATGATG 1466  
DB 1437 TCAACCAAGCAGCTGACTTATGTTTATTTTGGACATTTGATGATGATGATGATGATGATG 1496  
QY 1467 TCAATAAGGTTCAAAAGAGAGTACCTCTGTGACATCTTGGATGATGATGATGATGATGATG 1526  
DB 1497 TCAATAAGGTTCAAAAGAGAGTACCTCTGTGACATCTTGGATGATGATGATGATGATGATG 1556  
QY 1527 GTATAGATATACCAATGAAATTTCTGCTGCGAGGGGTGTAGAAATTTACTTTTTTGG 1586  
DB 1557 GTATAGATATACCAATGAAATTTCTGCTGCGAGGGGTGTAGAAATTTACTTTTTTGG 1616  
QY 1587 GTATATTTCTTATATATATTTATGATCATGCTGTGCTGAAATTTTATGATTTTGTGTTT 1646  
DB 1617 GTATATTTCTTATATATATTTATGATCATGCTGTGCTGAAATTTTATGATTTTGTGTTT 1676  
QY 1647 AATTAAGACTAACCAAACTTAATTAATAA 1678  
DB 1677 AATTAAGACTAACCAAACTTAATTAATAA 1708

**ABQ79518**

ID ABQ79518 standard; cDNA; 3145 BP.

XX AC  
XX AC  
XX AC  
XX DT  
XX DE  
XX KW  
KW KX  
XX OS  
XX XX  
XX FH  
FT FT  
XX CDS

25-NOV-2002 (first entry)

cDNA encoding a protein similar to human sorting nexin.  
  
Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;  
SSH 6; gene; human; nexin; ss.  
  
Homo sapiens.

Key Location/Qualifiers  
CDS 128..2968 /tag= a

WO200262839-A2.

PD PD  
PD PD  
PP PP  
PP PP  
PR PR  
PR PA  
PA XX  
PI PI  
PI PS  
DR DR  
DR PT  
PT PT  
PT PT  
PT PT  
PS PS  
XX XX

Damen MJAP, Cleutjens CBUM, Zaman GJR,  
WPI; 2002-643400/69.  
P-PADB; ABB81193.

Use of a polynucleotide differentially expressed in ruptured and stable atherosclerotic plaques as a maker for atherosclerosis, useful in treating, diagnosing or preventing atherosclerosis -  
  
Example 6; Page 35-37; 4app; English.

The invention relates to the use of a polynucleotide differentially expressed in ruptured and stable atherosclerotic plaques as a marker for atherosclerosis, where the polynucleotides can be selected from the sequences shown in ABQ79517-19. The polynucleotides are useful as a marker of atherosclerosis, which may be used:

- (i) in the diagnosis,
- prevention and treatment of atherosclerosis;
- (ii) as serum/plasma markers to screen patients at risk for plaque instability to evaluate the effects of other treatments;
- (iii) in the preparation of vector molecules for the expression of the encoded protein in host cells; and (iv) in the identification of functional targets or analogues of the gene. The polynucleotides, the encoded proteins or antibodies against the proteins may be used to target other therapeutics to an unstable plaque. Modulation of the expression of the polynucleotide can increase plaque stability and therefore inhibit the progression of atherosclerotic cardiovascular disease. Modulators may be used to prepare pharmaceuticals for atherosclerotic disorders. The present sequence represents a cDNA encoding a protein similar to human sorting nexin.

Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;

Query Match 65.1%; Score 1296; DB 24; Length 3145;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1446; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y Y  
Db Db  
Y Qy  
Db Db  
Y Qy  
Db Db  
Y Qy  
Sg Sg

147 GGAAACACAGAAAAGGGAATCTTTGGTAATCACAGAATAGGTAGCAATAAAG 206  
1602 GGAAACACAGAAAAGGGAATCTATTGGGAATCACAGAATAGGTAGCAATAAAG 1661  
207 GAGTATTCAAAGTCACCACAATGGAGGAGCATGTGCCTAANTATGTGTAGCTGAAG 266  
1662 GAGTATTCAAAGTACCCANAUGAGGAGACTATGTGCCTAANTATGTGTAGCTGAAG 1721  
267 GTGAAGATGTTTTATTGAAGAGGTATTGTGTGAATCGAAGATGATCTCCAGTGGAAG 326

QY 1407 TCACACGAGCTGACCTATGTTTATGACATTTGTGATACAGCACTGTTCCAGAC 1466  
DB 2862 TCACACGAGCTGACCTATGTTTATGACATTTGTGATACAGCACTGTTCCAGAC 2921  
QY 1467 TCAATAGGTACAAAAGGAGTTACCTCTGTGACATCTTGGATGTAAACACACTTGGATTG 1526  
DB 2922 TCAATAGGTACAAAAGGAGTTACCTCTGTGACATCTTGGATGTAAACACACTTGGATTG 2981  
QY 1527 GTATAGAAATACCCATTGAAATTTCTGTGCGAGGGTGTAGAAATTTACTTTTGG 1586  
DB 2982 GTATAGAAATACCCATTGAAATTTCTGTGCGAGGGTGTAGAAATTTACTTTTGG 3041  
QY 1587 GTATATTCT 1595  
DB 3042 GTATATTCT 3050

## RESULT 6

AAI59030  
ID AAI59030 standard; cDNA; 1551 BP.

AC AAI59030;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 1233.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM39874.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 1233; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX SQ Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;

Query Match 46.1%; Score 918; DB 22; Length 1551;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGAACACACAGAAAAGGGAGAAATCAATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 206

DB 177 GGAACACACAGAAAAGGGAGAAATCAATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 236

QY 207 GAGTATTCAAAAGTACCAATGAGGAGCTATGTTGCTTAATTTATGTTAGCTGAAG 266

DB 237 GAGTATTCAAAAGTACCAATGAGGAGCTATGTTGCTTAATTTATGTTAGCTGAAG 296

QY 267 GTGAGATGATTTTATTTGAAGAGGTATTTCTGTAATGGAAGATGATTCCTCAGTGGAG 326

DB 297 GTGAAGATGATTTTATTTGAAGAGGTATTTCTGTAATGGAAGATGATTCCTCAGTGGAG 356

QY 327 CTGTGACACACCTTAATACCTCCCGAAACCTCTGTCATCGAAATTTAGCATTTCCATATG 386

DB 357 CTGTGACACACCTTAATACCTCCCGAAACCTCTGTCATCGAAATTTAGCATTTCCATATG 416

QY 387 TAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGT 446

DB 417 TAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGT 476

QY 447 TTTGTATGATGTTGAAGAAATGATAGAGACGCTTGACACGCGCTGAACTTGGT 506

DB 477 TTTGTATGATGTTGAAGAAATGATAGAGACGCTTGACACGCGCTGAACTTGGT 536

QY 507 CTGTCTATAGAGATATCTTTGAATCTTATGTAATCAAACTAAACAGAAATTTCAATG 566

DB 537 CTGTCTATAGAGATATCTTTGAATCTTATGTAATCAAACTAAACAGAAATTTCAATG 596

QY 567 GTGCATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGCCCCCAAAATTTATGAAT 626

DB 597 GTGCATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGCCCCCAAAATTTATGAAT 656

QY 627 TCTTAAAGTCAAGAGAGGAGAGGTCCCAAGATATCTACAGAACTTCTCAGCATCCAG 686

DB 657 TCTTAAAGTCAAGAGAGGAGAGGTCCCAAGATATCTACAGAACTTCTCAGCATCCAG 716

QY 687 AACTGATTAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGTTGGGAAACACAAT 746

DB 717 AACTGATTAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGTTGGGAAACACAAT 776

QY 747 TTTCTGTATAGATATACTACAGATGTAATCTTGGGAAATTTAATAATCTGTTCTGGAA 806

DB 777 TTTCTGTATAGATATACTACAGATGTAATCTTGGGAAATTTAATAATCTGTTCTGGAA 836

QY 807 AACTAATGAAGAGAGGAGGTGAGCATTTGAACTTTTATCATGAATTTCAATTAATCTTT 866

DB 837 AACTAATGAAGAGAGGAGGTGAGCATTTGAACTTTTATCATGAATTTCAATTAATCTTT 896

QY 867 GTGAGTCTCCAAAGCTTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACTTCAG 926

DB 897 GTGAGTCTCCAAAGCTTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACTTCAG 956

QY 927 AAAACACACAGAACTTTTCAATGATCTGTTTAAAAATAATGAAACCGTCTGAAATA 986

DB 957 AAAACACACAGAACTTTTCAATGATCTGTTTAAAAATAATGAAACCGTCTGAAATA 1016

QY 987 CAGAGAGAAACAAAATCAGAAATTTATTTATGAGGATGATGACTGTAGAGAGTCTATG 1046

Db 1017 CACAGAGAAACGAAATCAGAAATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATG 1076

Qy 1047 ATTACCTGATGTATGTAG 1064

Db 1077 ATTACCTGATGTATGTAG 1094

RESULT 7

AAZ17379

ID AAZ17379 standard; cDNA; 779 BP.

XX AAC

XX ATC

XX DT

XX 12-OCT-1999 (first entry)

XX DE

XX Human gene expression product cDNA sequence SEQ ID NO:4852.

XX KW Human; gene; gene expression product; diagnosis; therapy; probe;

XX KW detection; mapping; tissue typing; profiling; forensic; cancer;

XX KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX OS Homo sapiens.

XX KW WO9938972-A2.

XX PN

XX OS 05-AUG-1999.

XX PD

XX PP 28-JAN-1999; 99WO-US01619.

XX PR 03-APR-1998; 98US-0080666.

XX PR 28-JAN-1998; 98US-0072910.

XX PR 24-FEB-1998; 98US-0075954.

XX PR 31-MAR-1998; 98US-0080114.

XX PR 03-APR-1998; 98US-0080515.

XX PA

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

XX PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

XX PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

XX PI Lanson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI, 1999-494092/41.

XX Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1; Page 2307; 2479pp; English.

XX The present invention describes a library of human polynucleotides

XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is

XX a method of detecting differentially expressed genes correlated with the

XX cancerous state of a mammalian cell, comprising detecting at least one

XX differentially expressed gene product in a test sample from a cell

XX suspected of being cancerous, where the gene product is encoded by one

XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The

XX polynucleotides can be used as a source of primers and probes, which can

XX be used for a variety of purpose, e.g. detection of expression levels,

XX mapping, tissue typing or profiling, forensics, genetic analysis and

XX detection of polymorphisms. Polypeptides encoded by the polynucleotides

XX can be used for raising antibodies for experimental, diagnostic and

XX therapeutic purposes. The polynucleotides may also be used to construct

XX arrays for diagnostics (which may be used to determine function of an

XX encoded protein); and to detect differences in expression levels between

XX two cells (e.g. to identify abnormal or diseased tissue in a human, to

XX identify a genetic predisposition or susceptibility to a disease such as

XX cancer). The polynucleotides of the invention are especially used in the

XX diagnosis, prognosis and management of colorectal cancer, breast cancer,

XX and lung cancer. The polynucleotides can also be used to screen for

XX peptide analogues and antagonists.

PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234059P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235113P.  
PR 26-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237605P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
PR 01-NOV-2000; 2000US-245084P.  
PA (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX Claim 1; SEQ ID 1893; 44pp; English.  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC esophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumour.  
SQ Sequence 451 BP; 179 A; 73 C; 52 G; 146 T; 1 other;

Query Match 19.3%; Score 385; DB 24; Length 451;  
Best Local Similarity 100.0%; Pred. No. 6.9e-165;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1597 ATATATATATGACATCGCTGCTGAAATTTAGTATATTTTGTATTAAGACT 1656  
DB 385 ATATATATATGACATCGCTGCTGAAATTTAGTATATTTTGTATTAAGACT 326  
QY 1657 AACACAACTTAATGATTAAGAGTGAATGATCTCATAGTCTTTTCATTGCTAGCTGTGA 1716  
DB 325 AACACAACTTAATGATTAAGAGTGAATGATCTCATAGTCTTTTCATTGCTAGCTGTGA 266  
QY 1717 TCCAAATTTTATAGAACATAAGTCACTTCTGTTATGTCATTTTAAAGAGAAATTCAT 1776  
DB 265 TCCAAATTTTATAGAACATAAGTCACTTCTGTTATGTCATTTTAAAGAGAAATTCAT 206  
QY 1777 AATGATGTTATGCGCAACAGATAAGACTGATAAAGTTCGTTATGTTATAGCTTTGAAATA 1836  
DB 205 AATGATGTTATGCGCAACAGATAAGACTGATAAAGTTCGTTATGTTATAGCTTTGAAATA 146  
QY 1837 ATTATGCTAGTAGTGGAGAACAGGATAAGATCTGATTTTCTAGAGTTAATATATTTT 1896  
DB 145 ATTATGCTAGTAGTGGAGAACAGGATAAGATCTGATTTTCTAGAGTTAATATATTTT 86  
QY 1897 AGTAGATTGTTTTCCTTTTATTTTATTTTGTACATAGTTAACTGTGTATCTATAATAA 1956  
DB 85 AGTAGATTGTTTTCCTTTTATTTTATTTTGTACATAGTTAACTGTGTATCTATAATAA 26  
QY 1957 GCATCCTATATAGAGTTTATTAAT 1981  
DB 25 GCATCCTATATAGAGTTTATTAAT 1

RESULT 9  
AAC02755  
ID AAC02755 standard; cDNA; 358 BP.  
XX AAC02755;  
XX  
DT 06-OCT-2000 (first entry)  
XX Human secreted protein 5' EST, SEQ ID NO: 2753.  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX (GEST) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
DR P-PSDB; AAG02749.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
PS Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs



CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-AT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC cDNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
SQ Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;

Query Match 17.9%; Score 357; DB 21; Length 358;  
Best Local Similarity 100.0%; Pred. No. 3.8e-152;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 ACTACAGATGAATCTTTGGGAAATATATAAATCTGTTCTGGAAACTAATGAAGA 819  
DB 1 ACTACAGATGAATCTTTGGGAAATATATAAATCTGTTCTGGAAACTAATGAAGA 60  
QY 820 GAAAGTCAGCTTTGGAACTTTTATCATGAATTTCAATTAATTTCTGAGTCTCCAAA 879  
DB 61 GAAAGTCAGCTTTGGAACTTTTATCATGAATTTCAATTAATTTCTGAGTCTCCAAA 120  
QY 880 GCTTAAACCAAGTAGACAGACTGACCAATTCAGCCCTACTTCAGAAAACACAGAA 939  
DB 121 GCTTAAACCAAGTAGACAGACTGACCAATTCAGCCCTACTTCAGAAAACACAGAA 180  
QY 940 GCTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATACAGAGAGAAGCA 999  
DB 181 GCTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATACAGAGAGAAGCA 240  
QY 1000 AAATCAGATTTATTTATGAGTGATGATCTAGAGAGTCTATGATTACCTGATGTA 1059  
DB 241 AAATCAGATTTATTTATGAGTGATGATCTAGAGAGTCTATGATTACCTGATGTA 300  
QY 1060 TGTAGACGGGTAGTTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAAC 1116  
DB 301 TGTAGACGGGTAGTTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAAC 357

RESULT 10  
ABN62287/c  
ID ABN62287 standard; cDNA; 580 BP.  
XX AC ABN62287;  
XX DT 28-JUN-2002 (first entry)  
XX DE Human cancer related polynucleotide SEQ ID NO 2254.  
XX KW Human; cytostatic; gene expression; gene mapping; tissue profiling;  
XX KW gene therapy; cancer; tumour; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200214500-A2.  
XX PD 21-FEB-2002.  
XX PF 16-AUG-2001; 2001WO-US25840.  
XX PF 16-AUG-2000; 2000US-226326P.  
XX PR (CHIR ) CHIRON CORP.  
XX PR (HYSE-) HYSEQ INC.  
XX PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;  
XX PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;  
XX DR WPI; 2002-241905/29.

XX New nucleic acid for producing a polypeptide, detecting differentially  
PT expressed genes correlated with a cancerous state of a mammalian cell,  
PT and inhibiting tumor growth -  
XX Claim 1; SEQ ID NO 2254; 883pp + Sequence Listing; English.  
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
CC with cytostatic activity. The polynucleotide is used to produce a  
CC polypeptide, to detect differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell and to inhibit tumour growth. The  
CC polynucleotide is used as a probe in mapping and tissue profiling. The  
CC encoded polypeptide and antibodies to the polypeptide can also be used  
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for  
CC gene therapy.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 580 BP; 242 A; 78 C; 94 G; 166 T; 0 other;

Query Match 3.9%; Score 78; DB 24; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.3e-25;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTATGAAACTCCAAAGTTATGTTAGCTATACCTTTAGTTTATCATCTTTCAAACTGTT 60  
DB 86 GTATGAAACTCCAAAGTTATGTTAGCTATACCTTTAGTTTATCATCTTTCAAACTGTT 27  
QY 61 TTTCTTTTCTTTTAAATAA 78  
DB 26 TTTCTTTTCTTTTAAATAA 9

RESULT 11  
AAH69552  
ID AAH69552 standard; cDNA; 311 BP.  
XX AC AAH69552;  
XX DT 19-SEP-2001 (first entry)  
XX DE Human cervical cancer marker nucleic acid 826.  
XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN WO200142467-A2.  
XX PD 14-JUN-2001.  
XX PF 08-DEC-2000; 2000WO-US33312.  
XX PF 08-DEC-1999; 99US-0169881.  
XX PF 21-DEC-1999; 99US-0171350.  
XX PF 14-MAR-2000; 2000US-0189315.  
XX PF 12-MAY-2000; 2000US-0203791.  
XX PF 09-JUN-2000; 2000US-0210600.  
XX PF 21-JUL-2000; 2000US-0220114.  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX DR WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
XX Claim 1; Page 243; 1051pp; English.  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with



GC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX  
SQ Sequence 311 BP; 84 A; 67 C; 63 G; 94 T; 3 other;

Query Match 3.8%; Score 76; DB 22; Length 311;  
Best Local Similarity 100.0%; Pred. No. 2.7e-24;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 1462  
DB 24 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 83

QY 1463 GAGCTCAATAAGGTAC 1478  
DB 84 GAGCTCAATAAGGTAC 99

## RESULT 12

AAH70887  
ID AAH70887 standard; cDNA; 313 BP.

XX  
AC AAH70887;

XX  
DT 19-SEP-2001 (first entry)

XX  
XX Human cervical cancer marker nucleic acid 2161.

XX  
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX  
OS Homo sapiens.

XX  
XX WO200142467-A2.

XX  
PN 14-JUN-2001.

XX  
XX 08-DEC-2000; 2000WO-US33312.

XX  
XX 08-DEC-1999; 99US-0169681.

XX  
XX 21-DEC-1999; 99US-0171350.

XX  
XX 14-MAR-2000; 2000US-0189315.

XX  
XX 12-MAY-2000; 2000US-0203791.

XX  
XX 09-JUN-2000; 2000US-0210600.

XX  
XX 21-JUL-2000; 2000US-0220114.

XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX  
XX Schlegel R, Deeds J, Berger A, Zhao X;

XX  
XX WPI; 2001-375006/39.

XX  
XX New isolated nucleic acid for diagnosing and treating cervical cancer

XX  
XX and for assessing and detecting compounds for treating the cancer -

XX  
XX Claim 1; Page 456; 1051pp; English.

XX  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX  
SQ Sequence 313 BP; 83 A; 70 C; 63 G; 97 T; 0 other;

Query Match 3.8%; Score 76; DB 22; Length 313;

Best Local Similarity 100.0%; Pred. No. 2.7e-24;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 1462  
DB 2 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 61

QY 1463 GAGCTCAATAAGGTAC 1478  
DB 62 GAGCTCAATAAGGTAC 77

## RESULT 13

AAH72445  
ID AAH72445 standard; cDNA; 299 BP.

XX  
AC AAH72445;

XX  
DT 19-SEP-2001 (first entry)

XX  
XX Human cervical cancer marker nucleic acid 3719.

XX  
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX  
OS Homo sapiens.

XX  
XX WO200142467-A2.

XX  
XX 14-JUN-2001.

XX  
XX 08-DEC-2000; 2000WO-US33312.

XX  
XX 08-DEC-1999; 99US-0169681.

XX  
XX 21-DEC-1999; 99US-0171350.

XX  
XX 14-MAR-2000; 2000US-0189315.

XX  
XX 12-MAY-2000; 2000US-0203791.

XX  
XX 09-JUN-2000; 2000US-0210600.

XX  
XX 21-JUL-2000; 2000US-0220114.

XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX  
XX Schlegel R, Deeds J, Berger A, Zhao X;

XX  
XX WPI; 2001-375006/39.

XX  
XX New isolated nucleic acid for diagnosing and treating cervical cancer

XX  
XX and for assessing and detecting compounds for treating the cancer -

XX  
XX Claim 1; Page 696; 1051pp; English.

XX  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX  
SQ Sequence 299 BP; 82 A; 64 C; 57 G; 96 T; 0 other;

Query Match 3.7%; Score 74; DB 22; Length 299;  
Best Local Similarity 100.0%; Pred. No. 2.2e-23;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1405 ACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCAGA 1464  
DB 1 ACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCAGA 60

QY 1465 GCTCAATAAGGTAC 1478  
DB 61 GCTCAATAAGGTAC 74

```
RESULT 14
AAH73007
ID AAH73007 standard; cDNA; 299 BP.
XX AC
XX AAH73007;
XX DT 19-SEP-2001 (first entry)
XX DE
XX Human cervical cancer marker nucleic acid 4281.
XX KW
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX OS
XX Homo sapiens.
XX PN WO200142467-A2.
XX PD 14-JUN-2001.
XX PF 08-DEC-2000; 2000WO-US33312.
XX PR 08-DEC-1999; 99US-0169681.
XX PR 21-DEC-1999; 99US-0171350.
XX PR 14-MAR-2000; 2000US-0189315.
XX PR 12-MAY-2000; 2000US-0203751.
XX PR 09-JUN-2000; 2000US-0210600.
XX PR 21-JUL-2000; 2000US-0220114.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI; 2001-375006/39.
XX PT New isolated nucleic acid for diagnosing and treating cervical cancer
XX PT and for assessing and detecting compounds for treating the cancer -
XX PS Claim 1; Page 941-942; 1051pp; English.
XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with
XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded
XX CC polypeptides are useful to assess if a patient is afflicted with
XX CC cervical cancer or has a pre-malignant condition; to monitor the
XX CC progression of cervical cancer or a premalignant condition in a patient;
XX CC and to select and/or assess the efficacy of a compound or therapy for
XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be
XX CC useful for gene therapy.
XX SQ Sequence 299 BP; 82 A; 64 C; 56 G; 95 T; 2 other;
Query Match 3.7%; Score 74; DB 22; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.2e-23;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1405 ACTCAACAGCAGCTGACTTATGTTTATTTGACATTGATACAGGAACTGTTCCAGA 1464
DB 1 ACTCAACAGCAGCTGACTTATGTTTATTTGACATTGATACAGGAACTGTTCCAGA 60
QY 1465 GCTCAATAAGGTAC 1478
DB 61 GCTCAATAAGGTAC 74
RESULT 15
ABN39970
ID ABN39970 standard; DNA; 60 BP.
XX AC
XX ABN39970;
XX DT 15-JUL-2002 (first entry)
XX DE
XX Human spliced transcript detection oligonucleotide SEQ ID NO:12718.
XX KW
XX splice variant; transcriptome; oligonucleotide library; ss.
XX OS
XX Homo sapiens.
XX PN WO200210449-A2.
XX PD 07-FEB-2002.
XX PF 20-JUL-2001; 2001WO-IB01903.
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX WPI; 2002-257383/30.
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes -
XX PS Example 1; SEQ ID 12718; 47pp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 60 BP; 22 A; 16 C; 9 G; 13 T; 0 other;
Query Match 3.0%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.5e-17;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 891 GTAGACCAGAACTGACCATTTCTCAGCCCTACTTCTCAGAAACAAACAGAGCTTTCAATG 950
DB 1 GTAGACCAGAACTGACCATTTCTCAGCCCTACTTCTCAGAAACAAACAGAGCTTTCAATG 60
Search completed: January 31, 2004, 16:17:08
Job time : 552 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 06:31:47 ; Search time 86 Seconds  
(without alignments)  
2386.548 Million cell updates/sec

Title: US-09-744-313A-1  
Perfect score: 465  
Sequence: 1 MYLIRCLIFRNTQKRGESF.....ELFPELNKVKQVTSVTSMW 465

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113513\_4617/app\_query.fasta\_1.647  
-DB=issued Patents NA -SUFFIX=rni -MINMATCH=0.1 -LOOCL=0  
-LOOPEXT=0 -UNITS=bits -FMF=fastap -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAFI=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 @CGN 1.1\_56 @runat\_30012004\_113513\_4617 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -YGAPOP=6  
-YGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	98.1	1716	4	US-09-620-312D-922
2	307	66.0	1551	4	US-09-620-312D-923
3	8	1.7	85	2	US-08-332-766A-44
4	8	1.7	740	4	US-09-342-681C-99
5	8	1.7	966	4	US-09-252-991A-9495
6	8	1.7	1239	4	US-09-252-991A-9666
7	8	1.7	1842	4	US-09-205-258-160
8	8	1.7	3043	3	US-09-008-271A-14
9	8	1.7	15231	3	US-09-128-155-16
10	8	1.7	176373	3	US-09-128-155-17
11	7	1.5	161	1	US-08-110-786A-4
12	7	1.5	161	1	US-08-110-786A-6

Sequence 43, Appl	1	US-08-435-684A-43	223	1.5	7
Sequence 43, Appl	2	US-08-934-877A-43	223	1.5	7
Sequence 43, Appl	3	US-08-871-678C-43	223	1.5	7
Sequence 991, App	4	US-09-252-991A-991	276	1.5	7
Sequence 998, App	5	US-09-313-294A-998	280	1.5	7
Sequence 3533, Ap	6	US-09-107-532A-3549	285	1.5	7
Sequence 3533, Ap	7	US-09-313-294A-3533	288	1.5	7
Sequence 5146, Ap	8	US-09-313-294A-5146	288	1.5	7
Sequence 2, Appli	9	US-08-175-388-2	321	1.5	7
Sequence 2, Appli	10	US-08-779-620-2	321	1.5	7
Sequence 2, Appli	11	US-08-818-726-2	321	1.5	7
Sequence 20, Appli	12	US-08-818-726-2	321	1.5	7
Sequence 1, Appli	13	US-09-018-584A-20	321	1.5	7
Sequence 8, Appli	14	US-09-326-039-1	365	1.5	7
Sequence 11, Appli	15	US-08-332-766A-8	370	1.5	7
Sequence 25, Appli	16	US-09-326-039-11	371	1.5	7
Sequence 561, App	17	US-08-332-766A-25	421	1.5	7
Sequence 19, Appl	18	US-09-221-017B-561	475	1.5	7
Sequence 194, App	19	US-08-332-766A-19	597	1.5	7
Sequence 13015, A	20	US-09-328-111-194	627	1.5	7
Sequence 37, Appl	21	US-09-252-991A-13015	657	1.5	7
Sequence 11270, A	22	US-08-529-878B-37	661	1.5	7
Sequence 6791, Ap	23	US-09-252-991A-11270	666	1.5	7
Sequence 982, App	24	US-09-252-991A-6791	702	1.5	7
Sequence 984, App	25	US-08-998-416-982	710	1.5	7
Sequence 6948, Ap	26	US-08-998-416-984	725	1.5	7
Sequence 6722, Ap	27	US-09-252-991A-6948	765	1.5	7
Sequence 101, App	28	US-09-252-991A-6722	780	1.5	7
Sequence 13, Appl	29	US-09-205-258-101	784	1.5	7
Sequence 27, Appl	30	US-08-981-803-13	797	1.5	7
Sequence 13, Appl	31	US-08-981-803-27	797	1.5	7
Sequence 27, Appl	32	US-08-983-440-13	797	1.5	7
Sequence 27, Appl	33	US-08-983-440-27	797	1.5	7
Sequence 13, Appl	34	US-09-367-895-13	797	1.5	7

ALIGNMENTS

RESULT 1

US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, tian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (71)..(1543)  
US-09-620-312D-922

## Alignment Scores:

Pred. No.: 0 Length: 1716  
Score: 456.00 Matches: 456  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.06% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIle 29  
DB 173 TTTCGGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGTACCAAAAT 232  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
DB 233 AAGAGGATATTCARAAAGTACCAATGAGGAGCTATGTTGCCCTAATATGGTGTAGCT 292  
QY 50 GluGlyGluAspPheIleGluGlyIleValValMetGluAspSerProVal 69  
DB 293 GAAGGTGAAGATGATTTTATGAAGAAGTATGTTGTAATGAAGATGATTTCTCCAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaIleThrPlyIleSerIlePro 89  
DB 353 GAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGCTGCATGGAATAATAGCATTCGA 412  
QY 90 TyrValAspPheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109  
DB 413 TATGTAGACTTTTGTGAGATCCCTCTCTGAAGAAGAGGAAAGAAAGAAATTCCT 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgArgAlaValGlyHisGluProGluHis 129  
DB 473 GTGTTTGTATGATGTTGAAGAATAATAGACAGCAGTGGACAGCCTGAACAT 532  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 533 TGTCTGTCTATAGAAGATATCTGAATCTATGATCTGAATCAAACTAACAGAAATTT 592  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProIleAsnTyr 169  
DB 593 CATGGTGCAATTCCTGATGCCAGCTCTCTTAAAGAGATCATTTGGCCCCCAAAATAT 652  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
DB 653 GAATCTTAAAGTCAAGAGGAGAGTCCAAAGAAATATCTACAGAACTTCTGCAGCAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
DB 713 CCAGAACTGAGTAATAGTCACTCTGCGAGACTTCTTCCCTTAATGTTGGGGAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229  
DB 773 CAATTTCTTGATAAGATATACCAAGTGAATCTTGGGAAATATATAAATCTGTTCT 832  
QY 230 GlyLysLeuMetLysGluLysGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
DB 833 GGAAACTAATGAAGAAGAAAGTCCAGCATTTGGAACTTTTATCATGAATTTTCATTAAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
DB 893 TCTTGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGCACTGACCAATCTCAGCCCTACT 952  
QY 270 SerGluAsnAsnLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
DB 953 TCAGAAACCAACAGAGCTTTTCAATGATCTGTTTAAATATATGCAACCGTGTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309

DB 1013 AATACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAGGAGTC 1072  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisHis 329  
DB 1073 TATGATTAATCTGATGATGATGAGGAGGAGTTTCCAGGTTCTCTGAGTGGCTTCAAT 1132  
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1133 CTCTTAATGGAACTCGAATCTCTTTAAACACCCCTGAAATGTATGATGATGATGAT 1192  
QY 350 LeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
DB 1193 CTTCACTGTAAACTAGAACAGCTATTTCAAGGAGCACCGTTTGTCTCACTCATACACTT 1252  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 1253 CTCAAGATGCTATATCTGTGAAACACACTGACCTCTCTCAAGATAGCAAAA 1312  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
DB 1313 GAGCAAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGAT 1372  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
DB 1373 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1432  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuPhePro 449  
DB 1433 GTACTCAACCAAGCAGCTGACTTATGTTTATTTGGAATGTTGATGATGATGATGAT 1492  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
DB 1493 GAGCTCAATAGGTACAAAGGAAGTTACCTCTGTGACATCTTGGATG 1540

## RESULT 2

US-09-620-312D-923  
Sequence 923, Application US/09620312D  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunging  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 923  
LENGTH: 1551  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:



QY 207 GAGTATTCAAAGTACCAATGAGGAGCTATGTTGCTTAATTATGTTAGTGTAGCTGAAG 266  
DB 237 GAGTATTCAAAGTACCAATGAGGAGCTATGTTGCTTAATTATGTTAGTGTAGCTGAAG 296  
QY 267 GTGAAGATGATTTTATGAAGAAGGTATGTTGTTAATGGAAGATGATTTCCAGTGGAGG 326  
DB 297 GTGAAGATGATTTTATGAAGAAGGTATGTTGTTAATGGAAGATGATTTCCAGTGGAGG 356  
QY 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTCGTCATGCGAATAATAGCATTCCATATG 386  
DB 357 CTGTGAGCACACCTTAATCTCCCGAAACCTTCGTCATGCGAATAATAGCATTCCATATG 416  
QY 387 TAGACATTTTGTGAGATCCCTCTCTGAAAGGAGGAGGAAAGAAAGAAATTCCTGTGT 446  
DB 417 TAGACATTTTGTGAGATCCCTCTCTGAAAGGAGGAGGAAAGAAAGAAATTCCTGTGT 476  
QY 447 TTTGTATTGATGTTGAAAGAAATGATAGAGACAGTGTGGAACAGAGCCCTGAAATGTT 506  
DB 477 TTTGTATTGATGTTGAAAGAAATGATAGAGACAGTGTGGAACAGAGCCCTGAAATGTT 536  
QY 507 CTGTCTATAGACATATCTTGAATCTTATCTACTTGAATCAAACTAAACAGATTTCTATG 566  
DB 537 CTGTCTATAGACATATCTTGAATCTTATCTACTTGAATCAAACTAAACAGATTTCTATG 596  
QY 567 GTGCATTTCTGATGCCAGCTTCCTCTTAAGAGGATCATTTGCCGCCAAAATTTATGAAT 626  
DB 597 GTGCATTTCTGATGCCAGCTTCCTCTTAAGAGGATCATTTGCCGCCAAAATTTATGAAT 656  
QY 627 TCTTAAGTCAAGAGGAGGAGTTCAGAAATATCTAAGAAATTTCTGAGCATCCAG 686  
DB 657 TCTTAAGTCAAGAGGAGGAGTTCAGAAATATCTAAGAAATTTCTGAGCATCCAG 716  
QY 687 AACTGATTAATAGTCACTTCGCGAGACTTTCTTCCCTTAATGTTGGGGGAAACACAT 746  
DB 717 AACTGATTAATAGTCACTTCGCGAGACTTTCTTCCCTTAATGTTGGGGGAAACACAT 776  
QY 747 TCTTGATAGACTACCAAGATGTAATCTTGGGAAATTTAATAATCTGTTCTCTGAA 806  
DB 777 TCTTGATAGACTACCAAGATGTAATCTTGGGAAATTTAATAATCTGTTCTCTGAA 836  
QY 807 AACTTAATGAAGAGAGAGTTCAGCATTTGGAACTTTTATCATGAATTTTCAATATCTT 866  
DB 837 AACTTAATGAAGAGAGAGTTCAGCATTTGGAACTTTTATCATGAATTTTCAATATCTT 896  
QY 867 GTGAGTCTCAAAGCCTTAACCAAGTAGACAGACTGACCACTTCTCAGCCCTACTTCAG 926  
DB 897 GTGAGTCTCAAAGCCTTAACCAAGTAGACAGACTGACCACTTCTCAGCCCTACTTCAG 956  
QY 927 ABAACACAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAAATA 986  
DB 957 ABAACACAGAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAAATA 1016  
QY 987 CAGAGAGAAAGCAAAATCAGATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATG 1046  
DB 1017 CAGAGAGAAAGCAAAATCAGATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATG 1076  
QY 1047 ATTACCTGATGATGATGAGGCGGTAGTTTCCAGGTTCTCTGACCTGGCTTCTATCTCT 1106  
DB 1077 ATTACCTGATGATGATGAGGCGGTAGTTTCCAGGTTCTCTGACCTGGCTTCTATCTCT 1136  
QY 1107 TAATGGGAATCGAATCTCTTTAAACACCCCTGGAATGATATGATTAATTAATCTTC 1166  
DB 1137 TAATGGGAATCGAATCTCTTTAAACACCCCTGGAATGATATGATTAATTAATCTTC 1196  
QY 1167 AGTGAATCTAGAACAGCTTTTTCAGGAGCAGGTTTGTCTCTCACTATACACTTCTCA 1226  
DB 1197 AGTGAATCTAGAACAGCTTTTTCAGGAGCAGGTTTGTCTCTCACTATACACTTCTCA 1256  
QY 1227 GAGATGCTATTTCTGTGAAACACCTGAACCTCGCTCTCTCAAGATAGCAAAAGGAG 1286  
DB 1257 GAGATGCTATTTCTGTGAAACACCTGAACCTCGCTCTCTCAAGATAGCAAAAGGAG 1316  
QY 1287 CAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATG 1346

DB 1317 CAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATG 1376  
QY 1347 GTGAAGAAACCAAGTATGAAGCAATCAGACTTCTGTTTGTAGGCTTACAGCAACCAAGTAC 1406  
DB 1377 GTGAAGAAACCAAGTATGAAGCAATCAGACTTCTGTTTGTAGGCTTACAGCAACCAAGTAC 1436  
QY 1407 TCACRAGCAGCTGACTTATGTTTATTGACATTTGACATTTGATACAGGAACTGTTTCCAGAGC 1466  
DB 1437 TCACRAGCAGCTGACTTATGTTTATTGACATTTGATACAGGAACTGTTTCCAGAGC 1496  
QY 1467 TCATTAAGGTACAAAGAAAGTATACCTCTGTGACATCTTGGATGTAAACACTTGGATTG 1526  
DB 1497 TCATTAAGGTACAAAGAAAGTATACCTCTGTGACATCTTGGATGTAAACACTTGGATTG 1556  
QY 1527 GTATAGAAATACCAATGAAATTTCTGCTGTGCGAGGCTGTAGAAATTTACTTTTTGG 1586  
DB 1557 GTATAGAAATACCAATGAAATTTCTGCTGTGCGAGGCTGTAGAAATTTACTTTTTGG 1616  
QY 1587 GTATATTTCTTATATATATATGATGATCGCTGCTGAAATTTTACTTTTTGTTTT 1646  
DB 1617 GTATATTTCTTATATATATATGATGATCGCTGCTGAAATTTTACTTTTTGTTTT 1676  
QY 1647 AATAAGACTAAACAAACTTAATGATTTAAA 1678  
DB 1677 AATAAGACTAAACAAACTTAATGATTTAAA 1708

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 5569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1378)  
US-09-620-312D-923

Query Match 46.1%; Score 918; DB 4; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-983-965-540  
; Sequence 540, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 540  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (339)  
; OTHER INFORMATION:  
; OTHER INFORMATION: Clone ID: 64-BOVMS1-017-Q1-E1-H12  
US-09-983-965-540

Alignment Scores:  
Pred. No.: 6.94e-73 Length: 402  
Score: 682.00 Matches: 127  
Percent Similarity: 99.25% Conservatives: 5  
Best Local Similarity: 95.49% Mismatches: 1  
Query Match: 27.99% Indels: 0  
DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-983-965-540 (1-402)

QY	205	AsnGlyGlyGluThrGlnPheLeuAspLysIleuPheProAspValLeuGlyLysIle	224
Db	3	AATGGTGGGAAACACAGTCTTCTGTAAGATACACAGACGTAAATCTTGGAAAT	62
QY	225	IleLysSerValProGlyLysLeuMetLysGluLysGlnHisLeuGluProPheIle	244
Db	63	ATAAATCTGTTCTGGAAACTAATGAAGAGAAAGTCAACATTGGACCTTTCATC	122
QY	245	MetAsnPheLeuSerCysGluSerProLysProLysProSerArgProGluLeuThr	264
Db	123	ATGAATTCATTAATCTTGTAATCTCCAAAGCCTAAACCGAGTAACACAGAACTGACC	182
QY	265	IleLeuSerProThrSerGluAsnAsnLysLeuPheAsnAspLeuPheLysAsnAsn	284
Db	183	ATTCAGCCCTACCTCAGAGAATAAAGAAAGCTTTTAATGATCTGTATAGAATAAT	242
QY	285	AlaAsnArgAlaGluAsnThrGluIleGlnAsnGlnAsnTyrPheMetGluValMet	304
Db	243	GCAACCGTCTGAGATACAGAAAGAGGCAAAATCAGAAATTTATGGAATGATG	302
QY	305	ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValPheGlnValPro	324
Db	303	ACTGTAGAAGAGCTATGATTACCTGATGATGTATGACAGGTGTTTCCAGATTCCT	362
QY	325	AspTrpLeuHisHisLeuLeuMetGlyThrArgIleLeu	337
Db	363	GACTGGCTTCATCATCTCTTAATGGGAACCGAATCCTC	401

RESULT 3

US-10-085-783A-25415  
; Sequence 25415, Application US/10085783A  
; Publication No. US2004003784A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.

APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 25415  
; LENGTH: 471  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (17)..(17)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (19)..(19)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (49)..(49)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (58)..(59)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-085-783A-25415

Alignment Scores:  
Pred. No.: 5.18e-69 Length: 471  
Score: 651.00 Matches: 131  
Percent Similarity: 94.24% Conservatives: 0  
Best Local Similarity: 94.24% Mismatches: 7  
Query Match: 26.71% Indels: 1  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-085-783A-25415 (1-471)

QY	328	HisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyr-ThrAs	347
Db	4	CATGACTGTAAANGANGAACTCGAATCTTTTAAACACACCTGNAATGTATNTTGA	63
QY	347	pTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeu	367
Db	64	TTACTATCTTCAGTGTAAACTAGAACAGCTATTTCAGAGACCGTTTGGTCTCACTCAT	123
QY	367	ethrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAsp	387
Db	124	AACACTTCTCAGAGATCTATATTCTGTGAAACACACTGAACCTCTCTCTCCAGATAA	183
QY	387	sGlnLysGlyValLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuVal	407
Db	184	GCAAAAGAGGCAAAACAGACTTTTGAAGAAATGTAATTAATTCAGATCTGTAGT	243
QY	407	lLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeu	427
Db	244	CAAGTGTATTGGTGAAGAAACCAAGTATGAAGCATCAGACCTCTGTGTGATGGCTTACA	303
QY	427	nGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValleGlnGlu	447
Db	304	GCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGTGATACAGGA	363